国 外 电 子 与 通 信 教 材 系 列

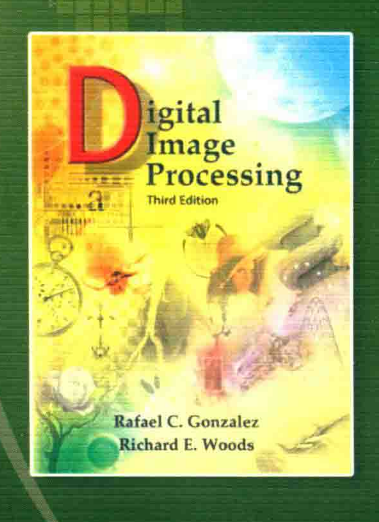


冈萨雷斯

数字图像处理

（第三版）

Digital Image Processing, Third Edition



[美] Rafael C. Gonzalez
Richard E. Woods 著
阮秋琦 阮宇智 等译



中国工信出版集团



電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

数字图像处理（第三版）

Digital Image Processing, Third Edition

在数字图像处理领域，本书作为主要教材已有30多年。就像由Gonzalez和Wintz编写的1977年版和1987年版、由Gonzalez和Woods编写的1992年版和2002年版那样，这一版也是我们用心为学生和教师准备的。基于对32个国家的134所高校的教师、学生和自学者的广泛调查，这一版已做了大量的修订与更新。新增内容与时俱进、可读性高，并给出了大量具有实际意义的例子。覆盖的图像处理领域包括灰度变换和空间滤波、频率域中的傅里叶变换和滤波、图像增强、图像复原、计算机断层、模糊集合、彩色图像处理、小波、图像和视频压缩、形态学处理、图像分割、图像描述、目标识别。

新 特 性

- 为使读者更早地接触基础内容，对介绍性概念进行了修订。
- 修订和更新了关于灰度变换、空间域相关、卷积及它们在空间滤波中的应用的讨论。
- 新增了关于模糊集合及其在图像处理中的应用的讨论。
- 扩充了关于离散傅里叶变换和频率域处理的内容。
- 扩充了关于计算机断层的内容。
- 对小波章节进行了较大的修订。
- 扩充了关于数据压缩的一章，包括新的压缩技术、数字视频压缩、压缩标准和水印处理。
- 扩充了关于形态学重建、灰度形态学和现代形态学算法的内容。
- 扩充了关于Marr-Hildreth边缘检测算法和坎尼边缘检测算法的内容。
- 扩充了关于图像阈值处理的内容。
- 给出了许多新例子，包括400余幅新图像和200多幅新图表。
- 扩充了习题集，包含了80多道新习题。
- 更新了参考文献。

相 关 图 书



ISBN 978-7-121-30540-5

本书网站 www.prenhall.com/gonzalezwoods

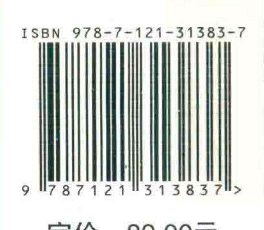
尽管本书完全自成体系，但近来全新改版的配套网站为读者提供了额外的支持，这些支持包括所选习题的答案、实验项目建议、课堂演示幻灯片等。教师手册可提供给采用本书作为教材的教师。





策划编辑：谭海平
责任编辑：谭海平
责任美编：孙焱津





定价：89.00元

国外电子与通信教材系列

数字图像处理

（第三版）

Digital Image Processing, Third Edition

［美］ Rafael C. Gonzalez
Richard E. Woods 著

阮秋琦　阮宇智　等译

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

本书是关于数字图像处理的经典著作，作者在对32个国家的134所院校和研究所的教师、学生及自学者进行广泛调查的基础上编写了第三版。除保留第二版的大部分主要内容外，还根据收集的建议从13个方面进行了修订，新增了400多幅图像、200多个图表和80多道习题，同时融入了近年来本科学领域的重要发展，使本书具有鲜明的特色与时效性。全书共分12章，包括绪论、数字图像基础、灰度变换与空间滤波、频域滤波、图像复原与重建、彩色图像处理、小波及多分辨率处理、图像压缩、形态学图像处理、图像分割、表现与描述、目标识别。

本书适用对象主要是从事信号与信息处理、计算机科学与技术、自动化、电子科学与技术、通信工程、信息工程、地球物理、生物工程、生物医学工程、物理、化学、医学、遥感等领域的大学教师和科技工作者、研究生、大学本科高年级学生及工程技术人员。

图书在版编目(CIP)数据

数字图像处理：第3版 / (美)拉斐尔·C·冈萨雷斯(Rafael C. Gonzalez)，(美)理查德·E·伍兹(Richard E. Woods)著；阮秋琦等译. —北京：电子工业出版社，2017.5
书名原文：Digital Image Processing, Third Edition
国外电子与通信教材系列
ISBN 978-7-121-31383-7

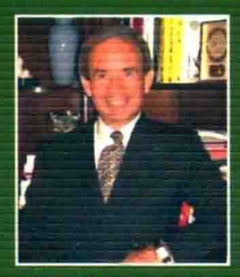
I. ①数… II. ①拉… ②理… ③阮… III. ①数字图像处理—高等学校—教材 IV. ①TN911.73

中国版本图书馆CIP数据核字(2017)第076988号

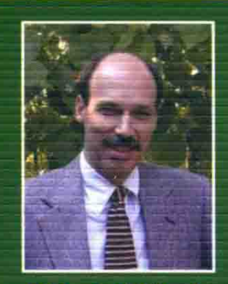
策划编辑：谭海平
责任编辑：谭海平
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路173信箱　　邮编　100036
开　　本：787×1092　1/16　印张：40.75　字数：1095千字　彩插：2
版　　次：2017年5月第1版（原著第3版）
印　　次：2017年5月第1次印刷
定　　价：89.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888，88258888。
质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。
服务热线：(010)88254552，tan02@phei.com.cn。

作者简介



Rafael C. Gonzalez



Richard E. Woods

译者简介



阮秋琦教授：1969年毕业于北方交通大学并留校任教。1981年于北方交通大学研究生毕业获工学硕士学位。1987年1月至1990年5月赴美国匹兹堡大学及辛辛纳提大学访问进修，主修图像处理和计算机视觉科学。1994年、1996、1997年再次赴美国任客座教授。现为教授、博士生导师、信息科学研究所所长、部级重点实验室主任。曾任通信与控制工程系主任、电子信息工程学院副院长、院长，计算机与信息技术学院院长。



图6.1

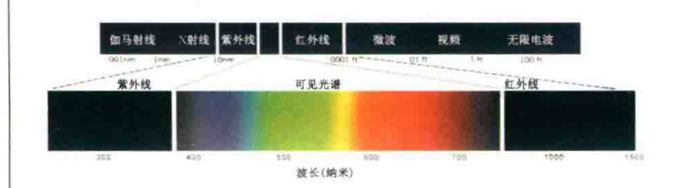


图6.2



图 6.4

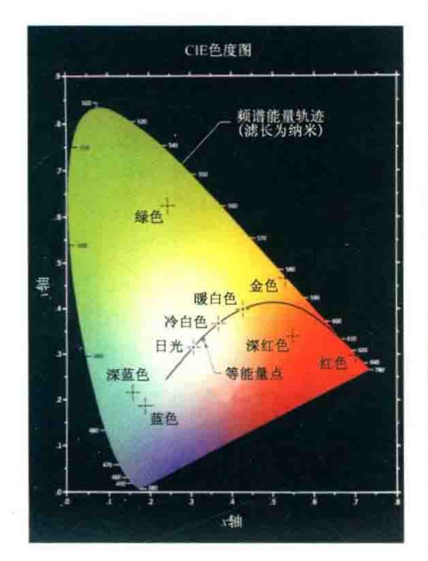


图 6.5

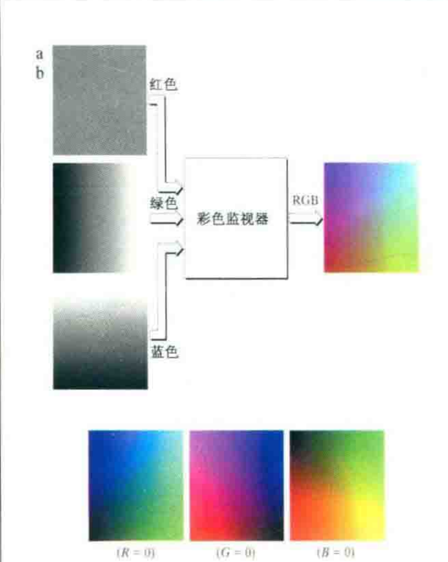


图 6.9

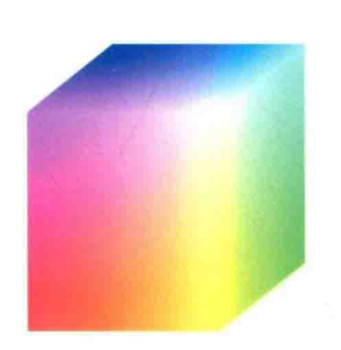


图 6.8



图 6.10

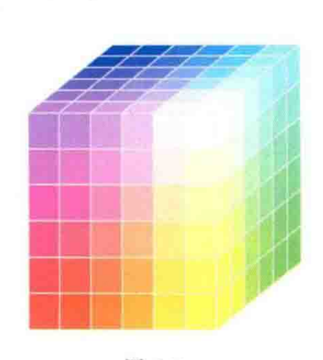


图 6.11

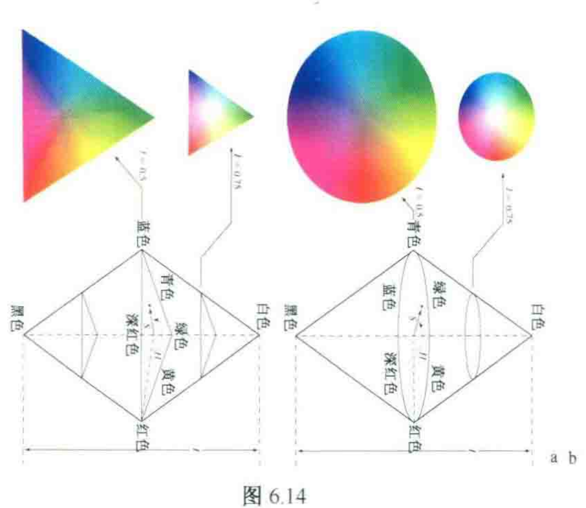


图 6.14

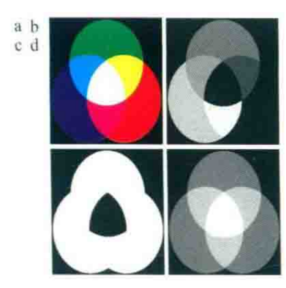


图 6.16

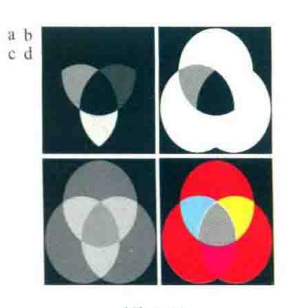


图 6.17

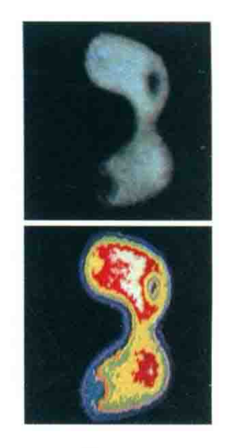


图 6.20

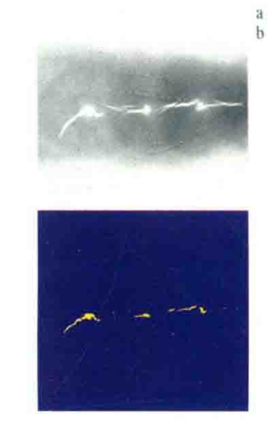


图 6.21

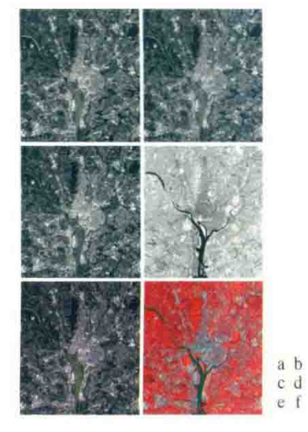


图 6.27

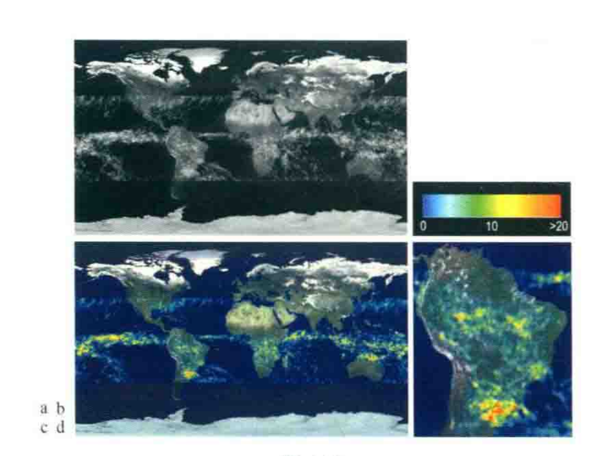


图 6.22

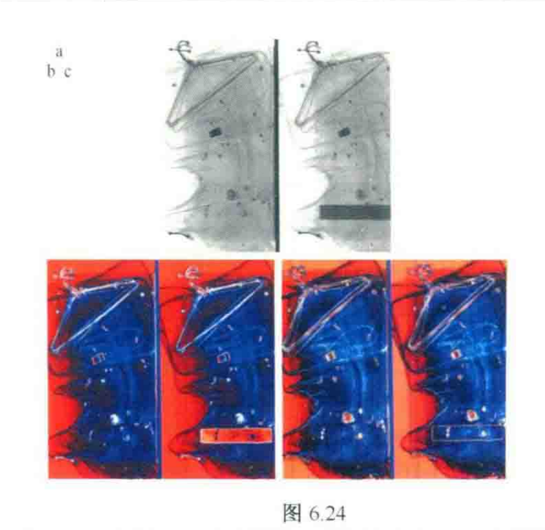


图 6.24

图 6.28

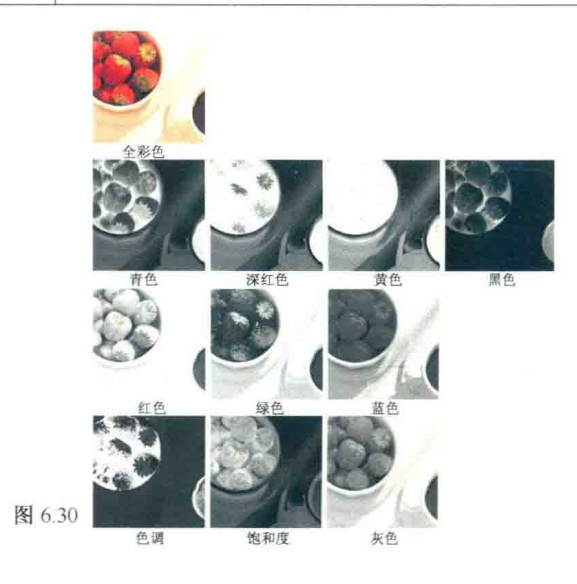


图 6.30



图 6.31

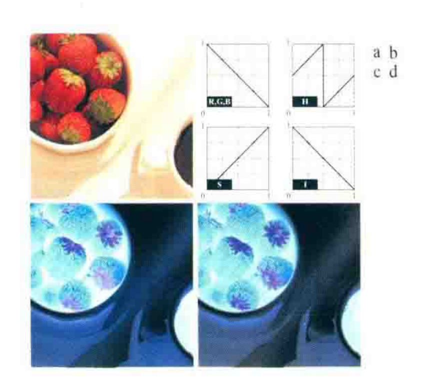


图 6.33

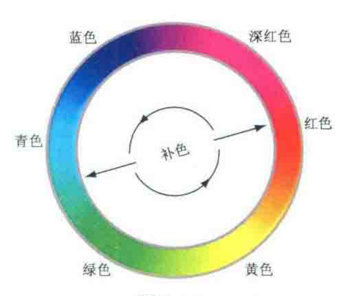


图 6.33

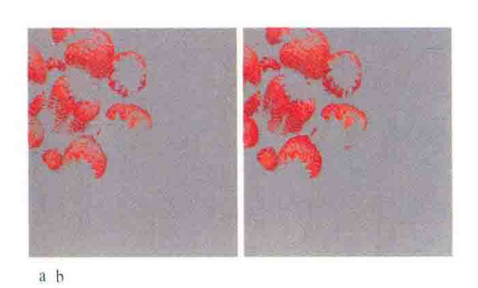


图 6.34



图 6.35



图 6.36

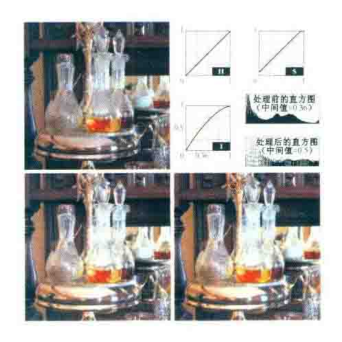


图 6.37



图 6.38



图 6.40



图 6.41

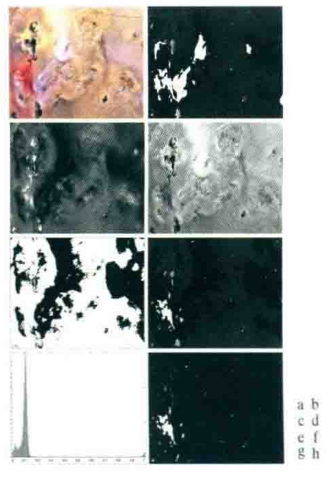


图 6.42



图 6.44

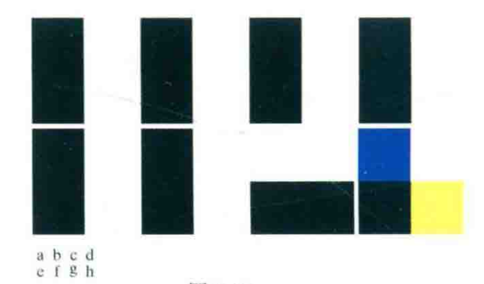


图 6.45

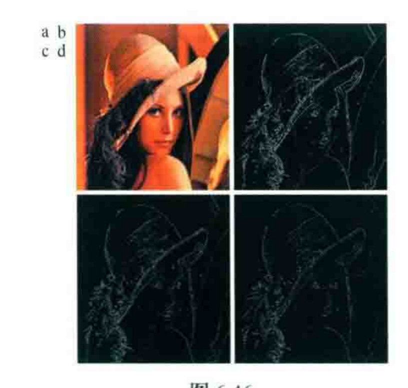


图 6.46



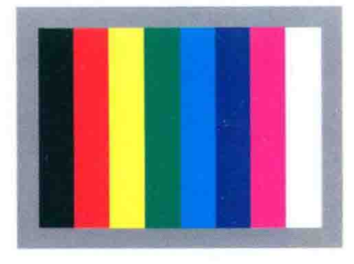
图 6.48



图 6.50



图 6.51



题 6.6图

序

2001 年 7 月间，电子工业出版社的领导同志邀请各高校十几位通信领域方面的老师，商量引进国外教材问题。与会同志对出版社提出的计划十分赞同，大家认为，这对我国通信事业、特别是对高等院校通信学科的教学工作会很有好处。

教材建设是高校教学建设的主要内容之一。编写、出版一本好的教材，意味着开设了一门好的课程，甚至可能预示着一个崭新学科的诞生。20 世纪 40 年代 MIT 林肯实验室出版的一套 28 本雷达丛书，对近代电子学科、特别是对雷达技术的推动作用，就是一个很好的例子。

我国领导部门对教材建设一直非常重视。20 世纪 80 年代，在原教委教材编审委员会的领导下，汇集了高等院校几百位富有教学经验的专家，编写、出版了一大批教材； 很多院校还根据学校的特点和需要，陆续编写了大量的讲义和参考书。这些教材对高校的教学工作发挥了极好的作用。近年来，随着教学改革不断深入和科学技术的飞速进步，有的教材内容已比较陈旧、落后，难以适应教学的要求，特别是在电子学和通信技术发展神速、可以讲是日新月异的今天，如何适应这种情况，更是一个必须认真考虑的问题。解决这个问题，除了依靠高校的老师和专家撰写新的符合要求的教科书外，引进和出版一些国外优秀电子与通信教材，尤其是有选择地引进一批英文原版教材，是会有好处的。

一年多来，电子工业出版社为此做了很多工作。他们成立了一个“国外电子与通信教材系列”项目组，选派了富有经验的业务骨干负责有关工作，收集了 230 余种通信教材和参考书的详细资料，调来了 100 余种原版教材样书，依靠由 20 余位专家组成的出版委员会，从中精选了 40 多种，内容丰富，覆盖了电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等方面，既可作为通信专业本科生和研究生的教学用书，也可作为有关专业人员的参考材料。此外，这批教材，有的翻译为中文，还有部分教材直接影印出版，以供教师用英语直接授课。希望这些教材的引进和出版对高校通信教学和教材改革能起一定作用。

在这里，我还要感谢参加工作的各位教授、专家、老师与参加翻译、编辑和出版的同志们。各位专家认真负责、严谨细致、不辞辛劳、不怕琐碎和精益求精的态度，充分体现了中国教育工作者和出版工作者的良好美德。

随着我国经济建设的发展和科学技术的不断进步，对高校教学工作会不断提出新的要求和希望。我想，无论如何，要做好引进国外教材的工作，一定要联系我国的实际。教材和学术专著不同，既要注意科学性、学术性，也要重视可读性，要深入浅出，便于读者自学； 引进的教材要适应高校教学改革的需要，针对目前一些教材内容较为陈旧的问题，有目的地引进一些先进的和正在发展中的交叉学科的参考书； 要与国内出版的教材相配套，安排好出版英文原版教材和翻译教材的比例。我们努力使这套教材能尽量满足上述要求，希望它们能放在学生们的课桌上，发挥一定的作用。

最后，预祝“国外电子与通信教材系列”项目取得成功，为我国电子与通信教学和通信产业的发展培土施肥。也恳切希望读者能对这些书籍的不足之处、特别是翻译中存在的问题，提出意见和建议，以便再版时更正。

吴佑寿

中国工程院院士、清华大学教授

“国外电子与通信教材系列”出版委员会主任

出版说明

进入 21 世纪以来，我国信息产业在生产和科研方面都大大加快了发展速度，并已成为国民经济发展的支柱产业之一。但是，与世界上其他信息产业发达的国家相比，我国在技术开发、教育培训等方面都还存在着较大的差距。特别是在加入 WTO 后的今天，我国信息产业面临着国外竞争对手的严峻挑战。

作为我国信息产业的专业科技出版社，我们始终关注着全球电子信息技术的发展方向，始终把引进国外优秀电子与通信信息技术教材和专业书籍放在我们工作的重要位置上。在2000年至2001年间，我社先后从世界著名出版公司引进出版了40余种教材，形成了一套“国外计算机科学教材系列”，在全国高校以及科研部门中受到了欢迎和好评，得到了计算机领域的广大教师与科研工作者的充分肯定。

引进和出版一些国外优秀电子与通信教材，尤其是有选择地引进一批英文原版教材，将有助于我国信息产业培养具有国际竞争能力的技术人才，也将有助于我国国内在电子与通信教学工作中掌握和跟踪国际发展水平。根据国内信息产业的现状、教育部《关于“十五”期间普通高等教育教材建设与改革的意见》的指示精神以及高等院校老师们反映的各种意见，我们决定引进“国外电子与通信教材系列”，并随后开展了大量准备工作。此次引进的国外电子与通信教材均来自国际著名出版商，其中影印教材约占一半。教材内容涉及的学科方向包括电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等，其中既有本科专业课程教材，也有研究生课程教材，以适应不同院系、不同专业、不同层次的师生对教材的需求，广大师生可自由选择和自由组合使用。我们还将与国外出版商一起，陆续推出一些教材的教学支持资料，为授课教师提供帮助。

此外，“国外电子与通信教材系列”的引进和出版工作得到了教育部高等教育司的大力支持和帮助，其中的部分引进教材已通过“教育部高等学校电子信息科学与工程类专业教学指导委员会”的审核，并得到教育部高等教育司的批准，纳入了“教育部高等教育司推荐——国外优秀信息科学与技术系列教学用书”。

为做好该系列教材的翻译工作，我们聘请了清华大学、北京大学、北京邮电大学、南京邮电大学、东南大学、西安交通大学、天津大学、西安电子科技大学、电子科技大学、中山大学、哈尔滨工业大学、西南交通大学等著名高校的教授和骨干教师参与教材的翻译和审校工作。许多教授在国内电子与通信专业领域享有较高的声望，具有丰富的教学经验，他们的渊博学识从根本上保证了教材的翻译质量和专业学术方面的严格与准确。我们在此对他们的辛勤工作与贡献表示衷心的感谢。此外，对于编辑的选择，我们达到了专业对口； 对于从英文原书中发现的错误，我们通过与作者联络、从网上下载勘误表等方式，逐一进行了修订； 同时，我们对审校、排版、印制质量进行了严格把关。

今后，我们将进一步加强同各高校教师的密切关系，努力引进更多的国外优秀教材和教学参考书，为我国电子与通信教材达到世界先进水平而努力。由于我们对国内外电子与通信教育的发展仍存在一些认识上的不足，在选题、翻译、出版等方面的工作中还有许多需要改进的地方，恳请广大师生和读者提出批评及建议。

电子工业出版社

教材出版委员会

译 者 序

数字图像处理起源于20世纪20年代，当时通过海底电缆从英国的伦敦到美国的纽约采用数字压缩技术传输了第一张数字照片。此后，由于遥感等领域的应用，图像处理技术逐步受到关注并得到相应的发展。由于技术手段的限制，图像处理科学与技术的发展相当缓慢，直到第三代计算机问世后，数字图像处理才开始迅速发展并得到普遍应用。CT的发明、应用及获得倍受科技界瞩目的诺贝尔奖，使得图像处理技术大放异彩。目前数字图像处理科学已成为工程学、计算机科学、信息科学、统计学、物理学、化学、生物学、医学甚至社会科学等领域学习和研究的对象。随着信息高速公路、数字地球概念的提出及互联网的广泛应用，图像处理技术的需求与日俱增。其中，图像信息以其信息量大、传输速度快、作用距离远等一系列优点，成为人类获取信息的重要来源及利用信息的重要手段；因此，图像处理科学与技术逐步向其他学科领域渗透并为其他学科利用是理所当然的。图像处理科学又是一门与国计民生紧密相关的应用科学，它已给人类带来了巨大的经济效益和社会效益，不久的将来它不仅在理论上会有更深入的发展，在应用上亦是科学研究、社会生产乃至人类生活中不可缺少的有力工具。它的发展及应用与我国的现代化建设联系之密切、影响之深远是不可估量的。在信息社会中，图像处理科学无论是在理论上还是在实践中，都有着巨大的潜力。

本书第一版于 1977 年问世，它是作者在为大学高年级学生和研究生编写的讲义的基础上整理而成的，全书只有七章。该书深入浅出，图文并茂，概念清楚，通俗易懂，是很受欢迎的教科书。特别是对非英语国家的学生和科技工作者，阅读原文也会觉得赏心悦目，朗朗上口，给读者留下了深刻的印象。我想这是该书被广泛用做教材的根本所在。本书是该书的第三版，这一版除保留了该书的传统风格外，在内容上进行了大幅度的修订与重组，作者根据反馈意见从 13 个方面对本书进行了修订，新增了 400 余幅图像、200 多幅图表及 80 多道习题，同时融入了近年来本科学领域的重要进展，使得本书更加充实与全面。

本书共分 12 章，分别为绪论、数字图像基础、灰度变换与空间滤波、频率域滤波、图像复原与重建、彩色图像处理、小波与多分辨率处理、图像压缩、形态学图像处理、图像分割、表示与描述、目标识别。

为了统一全书的语言风格，参加翻译的人员较少。全书第 1 章至第 7 章及前言、目录、致谢、关于作者等由阮秋琦翻译，其余内容均由阮宇智同志翻译；全书由阮秋琦统一整理与审校。由于时间仓促，难以达到“信、达、雅”的高标准，退而求其次，尽量做到译文准确，风格统一。

本书在翻译中得了许多同学的帮助，对此，译者深表感谢。由于译者水平所限，书中一定会有许多错误和不当之处，恳切地希望读者提出宝贵的建议和批评。

译　者

前　言

When something can be read without effort, great effort has gone into its writing.

Enrique Jardiel Poncela

第三版是本书的一次重要修订。如同由 Gonzalez 和 Wintz 编写的 1977 年版和 1988 年版，以及由 Gonzalez 和 Woods 编写的 1992 年版和 2002 年版那样，这一版同样是为学生和教师考虑而准备的。本书的主要目的仍是介绍数字图像处理的基本概念和方法，并为读者在该领域进一步学习和研究打下坚实的基础。为实现这一目的，我们仍将重点放在基础知识和普通应用上。本书要求读者的数学知识具备大学本科高年级和研究生一年级的水平，即需要掌握数学分析、向量、矩阵、概率、统计、线性系统和计算机编程方面的基本知识。本书的 Web 网站为读者提供了所需背景知识的回顾指南。

本书在数字图像处理领域处于引领地位 30 多年的主要原因是，我们对读者不断变化的教育需求给予了极大的关注。第三版是在我们广泛调查的基础上编写的，这些调查涉及 32 个国家的 134 所高校和研究机构的教师、学生与自学者。根据调查的反馈情况，本书做了如下修订：

- 更早、更全面地介绍了图像处理中所用的数学工具。
- 扩充说明了直方图处理技术。
- 逐步叙述了复杂的算法。
- 扩充说明了空间相关和卷积的内容。
- 介绍了模糊集合理论及其在图像处理中的应用。
- 修订了频率域处理的内容，从基本原理开始，说明了如何从数据取样得出离散傅里叶变换。
- 覆盖了关于计算机断层（CT）的内容。
- 清楚地叙述了小波的基本概念。
- 修订了关于数据压缩的内容，包含了更多的视频压缩技术、标准和水印。
- 扩充了形态学的章节，包含了形态学重建的内容，修订了灰度形态学的内容。
- 扩充了图像分割的内容，包含了更先进的技术，如坎尼算法的边缘检测技术，更全面地探讨了图像的阈值处理。
- 更新了图像表示与描述章节的内容。
- 精简了关于结构目标识别的内容。

第三版中的新内容和重新组织的内容试图在论述的严密性、描述的清晰性和市场调查的反馈之间保持平衡，同时尽量将篇幅控制在合理的范围内。第三版的主要改动如下。

第 1 章：更新了图片，并根据后续章节的变化重写了正文部分。

第2章：本章约修订了 50%的内容，包含了新图像和更清晰的说明。主要修订包括：新增了关于图像内插的一节，以及综述本书所用主要数学工具的一节。此前分散在全书中的大量图像处理应用现在整合到了第 2 章中。例如，我们把图像平均和图像相减移到了这一章。这遵循了我们在第二版中就开始的做法，

即在讨论中尽可能把许多应用前移，以便更好地引导读者。学完重新组织的第 2 章后，读者可基本了解数字图像加工和处理的方式。本章是编写本书其余章节的坚实基础。

第 3 章：本章的主要修订是，详细探讨了空间相关、卷积及利用空间模板对图像滤波的应用。我们在进行市场调查时发现了读者普遍关心的一个问题，即要求用大量实例来说明直方图均衡化和规定化，对这一问题的回应是，我们增加了一些说明这些处理工具的例子。模糊集合及其在图像处理中的应用也是调查反馈普遍要求的内容，因此本章中纳入了模糊集合理论基础及其在灰度变换与空间滤波两种图像处理中的主要应用。

第 4 章：过去 4 年，读者抱怨得最多的是对第一版和第二版的第 4 章所做的更改。当时做出这些改变时，目的是为了简化傅里叶变换和频率域的表述。显然，我们走得太远，因为读者抱怨新内容太浅。第三版纠正了这一问题。现在的内容从连续变量的傅里叶变换开始，再用取样和卷积的基本概念进一步推导了离散傅里叶变换。这种变化的优点是，更直观地引入了取样定理。然后，我们将一维情形推广到了二维情形，并给出了说明数字图像取样效果的一些例子。再后，我们介绍了二维离散傅里叶变换，推导和总结了一些重要性质。这些概念是频率域滤波的基础。最后，我们讨论了问题实现，如变换分解和快速傅里叶变换算法的推导。学完本章后，读者可掌握从一维函数的取样到离散傅里叶变换基础的清晰推导过程，以及其在数字图像处理中的某些重要应用。

第 5 章：本章增加了关于从投影重建图像的一节，重点在于计算机断层(CT)。CT 的内容用投影重建图像的基本原理和实践中所用成像模型的例子开始。然后，推导了雷登变换和傅里叶切片定理，并以它们为基础清楚地说明了滤波反投影的概念。讨论了平行光束和扇形光束重建，并用一些例子进行了说明。这些内容较老，但对本书是重要的补充。

第6章：本章只做了符号表示的澄清和更正，未增加新内容。

第7章：读者反馈从前一章过渡到小波对初学者来说较为困难，因此我们重写了一些基础内容。

第8章：为使内容跟上发展形势，本章已完全重写。新编码技术内容扩展到了视频，修订了标准，介绍了图像水印处理。这种新的编排方式更易于学生掌握。

第 9 章：本章的主要变化是包含了形态学重建的新内容，修订了灰度级形态学的内容，并详细介绍了二值图像和灰度级图像的形态学重建，以便学生开发出更复杂、更有用的形态学算法。

第10章：本章做了大量修订。组织方式与之前相同，但新增了关于分割技术的内容，详细讨论并说明了边缘模型及其性质。介绍先进的边缘检测技术时，包含了 Marr-Hildreth 边缘检测器和坎尼边缘检测器。重写了关于阈值处理的一节，包含了较为流行的 Otsu 方法，这种方法易于理解和实现，且应用广泛。关于贝叶斯方法的内容则移到第 12 章介绍，那时还会详细探讨贝叶斯决策规则。此外，还讨论了如何利用边缘信息来改进阈值处理，并给出了自适应阈值处理的新例子。关于形态学分水岭和运动分割应用的内容，除阐述更为清楚外，基本上未做改动。

第11章：本章的主要变化是边界跟踪算法，详细推导了用最小周长多边形拟合数字边界的算法，增加了用于纹理描述的共生矩阵内容。与 11.4 节的所有例子一样，11.2 节和 11.3 节中的许多例子都是新的。

第12章：本章的变化是，新增了关于相关匹配的内容，以及用贝叶斯分类器识别多光谱图像中感兴趣区的新例子。结构分类方面的章节只限于讨论串匹配。

以上修订，导致了本书新增了 400 多幅图像、200 多幅图表和 80 多道习题。书中适当的位置给出了复杂处理过程的逐步算法。同时，更新了参考文献。

本书的 Web 网站在第二版发行期间就已完成，并取得了很大的成功，每月的访问量都在20 000以

上。相应于第三版本，我们重新设计和升级了这一网站。关于该网站的详细功能和内容，请读者参阅后面的“本书网站”和“致谢”部分。

第三版反映了 2002 年以来读者不断变化的需求。自1977年首次出版以来，本书在全球范围内被读者广泛接受的原因之一是，本书一直强调基本概念，包括试图提供尽快引出知识主体的稳定方法。我们遵循相同的原则编写了本书的第三版。

Rafael C. Gonzalez

Richard E. Woods

致　谢

感谢高校、业界和政府中为本书做出贡献的诸多同仁，你们的贡献非常重要。特别感谢同事 Mongi A. Abidi、Steven L. Eddins、Yongmin Kim、Bryan Morse、Andrew Oldroyd、Ali M. Reza、Edgardo Felipe Riveron、Jose Ruiz Shulcloper 和 Cameron H. G. Wright，感谢你们在表达方式和覆盖范围方面所提出的建议。

撰写第三版期间，许多个人和组织为我们提供了有价值的帮助。特别感谢 Mathworks 公司的 Courtney Esposito 和 Naomi Fernandes，感谢你们为我们提供了 MATLAB 软件与支持，这对创建或阐述第三版中的许多实例和实验结果非常重要。第三版中的很多新图像都是通过个人努力得到的，在此向你们深表谢意，特别感谢 Serge Beucher、Melissa D. Binde、James Blankenship、Uwe Boos、Ernesto Bribiesca、Michael E. Casey、Michael W. Davidson、Susan L. Forsburg、Thomas R. Gest、Lalit Gupta、Daniel A. Hammer、Zhong He、Roger Heady、Juan A. Herrera、John M. Hudak、Michael Hurwitz、Chris J. Johannsen、Rhonda Knighton、Don P. Mitchell、Ashley Mohamed、A. Morris、Curtis C. Ober、Joseph E. Pascente、David. R. Pickens、Michael Robinson、Barrett A. Schaefer、Michael Shaffer、Pete Sites、Sally Stowe、Craig Watson、David K. Wehe 和 Robert A. West。还要感谢许多图题中引用的个人和企业，感谢你们允许我们使用这些图片。

特别感谢 Prentice Hall 公司的 Vince O'Brien、Rose Kernan、Scott Disanno、Michael McDonald、Joe Ruddick、Heather Scott 和 Alice Dworkin，感谢你们在本书制作中的创造性、帮助和耐心。

Rafael C. Gonzalez
Richard E. Woods

关 于 作 者

Rafael C. Gonzalez（拉斐尔·C·冈萨雷斯）

1965于美国迈阿密大学获电气工程学士学位；1967年和1970年于美国佛罗里达大学盖恩斯维尔分校分别获电气工程硕士学位和博士学位。1970年，加盟田纳西大学诺克斯维尔分校（UTK）电机和计算机工程系。1973年晋升为副教授，1978年晋升为教授，1984年被授予“杰出贡献教授”。1994年到1997年任系主任，现为UTK名誉教授。

冈萨雷斯是田纳西大学“图像与模式分析实验室”和“机器人与计算机视觉实验室”的创始人。还于1982年创立了感知公司，并任公司总裁直至1992年，期间的后三年全职受聘于其在1989年收购的Westinghouse公司。

在其引领下，感知公司在图像处理、计算机视觉和光盘存储技术等领域取得了很大的成功。前10年，公司推出了一系列创新产品，包括：计算机视觉系统，这是全球首台商用系统，可自动阅读行驶车辆的车牌；大规模图像处理和归档系统，美国海军在不同地点使用该系统来检测三叉II潜艇项目中导弹的火箭发动机；为苹果计算机设计的图像卡；万亿（10^{12}）字节光盘生产线。

他还是模式识别、图像处理和机器学习领域的企业和政府顾问。他在这些领域获得的荣誉包括：1977年获UTK工学院职员成就奖；1978年获UTK校长研究学者奖；1980年获Magnavox工程教授奖；1980年获M. E. Brooks杰出教授奖；1981年，成为田纳西大学的IBM教授并于1984年被授予杰出成就教授；1985年，被迈阿密大学授予杰出男毕业生奖；1986年，被授予Phi Kappa Phi学者奖；1992年，获田纳西大学Nathan W. Dougherty工程优秀奖。

工业界荣誉包括：1987年因田纳西的商业开发获IEEE杰出工程师奖；1988年因商业图像处理的优秀表现获Albert Rose Nat奖；1989年因在技术转化方面的优秀表现获Otto Wheeley奖；1989年获Coopers和Lybrand发起人年度奖；1992年获IEEE第3区杰出工程师奖；1993年因技术开发获自动成像协会国家奖。

冈萨雷斯在模式识别、图像处理和机器人领域发表或合作发表了100多篇论文，出版了两本专著及4本教材，被遍布全球的1000多所大学和研究机构采用。他被《全美名人传记》、《工程名人传记》、《世界名人传记》和10多个其他传记收录。他是两个美国专利的共同持有者，同时一直是*IEEE Transactions on Systems*、*Man and Cybernetics*和*International Journal of Computer and Information Sciences*的副主编。他是许多专业学会和名誉学会的会员，如Tau Beta Pi、Phi Kappa Phi、Eta Kappa Nu和Sigma Xi。他还是IEEE会士。

Richard E. Woods（理查德·E·伍兹）

理查德·E·伍兹，UTK电气工程系学士、硕士和博士。从业经历包括企业家、大学教师、咨询、企业管理和工业工程。最近创立了专门开发医用手持式计算机系统的MedData Interactive公司。他还是感知公司的创始人和副总裁，负责公司的定量图像分析和自动决策产品的开发工作。

在加入感知公司和 MedData 公司前，伍兹博士是田纳西大学电气工程和计算机科学专业的助理教授，再之前是 Union Carbide 公司的计算机应用工程师。作为顾问，他为许多空间和军事机构开发了大量专用数字处理器，这些机构包括 NASA、弹道导弹司令部和橡树岭国家实验室。

伍兹博士发表了大量有关数字信号处理方面的文章，并且是几个专业学会的成员，包括 Tau Beta Pi、Phi Kappa Phi 和 IEEE。1986 年，获得田纳西大学杰出工程毕业生称号。

谨以此书献给 Smmantha、Janice、David 和 Jonathan。

本 书 网 站

www.prenhall.com/gonzalezwoods
或镜像网站
www.imageprocessingplace.com

本书自成体系，但参考网站为许多重要的领域提供了有力的支持。
对于学生或自学者，网站内容包括：

- 概率、统计、向量和矩阵的回顾。
- 部分习题的完整解答。
- 计算机项目。
- 本书所讨论大部分主题的辅导内容。
- 本书中所用的全部图像。

对于教师，网站内容包括：

- 书中所有习题的解答及课堂和实验教学手册，这些内容对采用本书作为教材的教师免费。
- 课堂教学 PPT。
- 以前版本中删除的内容（PDF 格式）。
- 其他教学资源链接。

对于专业人员，网站中还包括一些特殊的主题：

- 商业网站链接。
- 部分新参考文献。
- 商业图像数据库链接。

图书出版后，为确保图书内容的更新，在需要添加新课题、数字图像和其他相关内容时，Web 网站是很方便的场所。虽然在出版本书时已做了仔细考虑，但网站对于发布任何印刷错误仍是很方便的场所。Web 网站在书中设计了如下图形：



Pearson

尊敬的老师：

您好！

为了确保您及时有效地申请培生整体教学资源，请您务必完整填写如下表格，加盖学院的公章后传真给我们，我们将会在 2-3 个工作日内为您处理。

请填写所需教辅的开课信息：

<table>
<tr><td>采用教材</td><td colspan="2"></td><td>□中文版 □英文版 □双语版</td></tr>
<tr><td>作　　者</td><td></td><td>出版社</td><td></td></tr>
<tr><td>版　　次</td><td></td><td>ISBN</td><td></td></tr>
<tr><td rowspan="2">课程时间</td><td>始于　　年　月　日</td><td>学生人数</td><td></td></tr>
<tr><td>止于　　年　月　日</td><td>学生年级</td><td>□专 科　　□本科 1/2 年级
□研究生　　□本科 3/4 年级</td></tr>
</table>

请填写您的个人信息：

<table>
<tr><td>学　　校</td><td colspan="3"></td></tr>
<tr><td>院系/专业</td><td colspan="3"></td></tr>
<tr><td>姓　　名</td><td></td><td>职　　称</td><td>□助教 □讲师 □副教授 □教授</td></tr>
<tr><td>通信地址/邮编</td><td colspan="3"></td></tr>
<tr><td>手　　机</td><td></td><td>电　　话</td><td></td></tr>
<tr><td>传　　真</td><td colspan="3"></td></tr>
<tr><td>official email(必填)
(eg:XXX@ruc.edu.cn)</td><td></td><td>email
(eg:XXX@163.com)</td><td></td></tr>
<tr><td colspan="4">是否愿意接受我们定期的新书讯息通知：　□是　　□否</td></tr>
</table>

系 / 院主任：__________________（签字）

（系 / 院办公室章）

_____年_____月_____日

资源介绍：

—教材、常规教辅（PPT、教师手册、题库等）资源：请访问www.pearsonhighered.com/educator；（免费）

—MyLabs/Mastering 系列在线平台：适合老师和学生共同使用；访问需要 Access Code；（付费）

100013　北京市东城区北三环东路 36 号环球贸易中心 D 座 1208 室
电话：（8610）57355003　　传真：（8610）58257961

Please send this form to:

目　录

第1章　绪论 …… 1
引言 …… 1
1.1　什么是数字图像处理 …… 1
1.2　数字图像处理的起源 …… 2
1.3　使用数字图像处理领域的实例 …… 4
1.3.1　伽马射线成像 …… 5
1.3.2　X射线成像 …… 5
1.3.3　紫外波段成像 …… 7
1.3.4　可见光及红外波段成像 …… 7
1.3.5　微波波段成像 …… 10
1.3.6　无线电波段成像 …… 10
1.3.7　使用其他成像方式的例子 …… 11
1.4　数字图像处理的基本步骤 …… 14
1.5　图像处理系统的组成 …… 15
小结 …… 17
参考文献 …… 17
第2章　数字图像基础 …… 20
引言 …… 20
2.1　视觉感知要素 …… 20
2.1.1　人眼的结构 …… 20
2.1.2　眼睛中图像的形成 …… 22
2.1.3　亮度适应和辨别 …… 22
2.2　光和电磁波谱 …… 24
2.3　图像感知和获取 …… 26
2.3.1　使用单个传感器获取图像 …… 27
2.3.2　使用条带传感器获取图像 …… 27
2.3.3　使用传感器阵列获取图像 …… 28
2.3.4　简单的图像形成模型 …… 28
2.4　图像取样和量化 …… 30
2.4.1　取样和量化的基本概念 …… 30
2.4.2　数字图像表示 …… 31
2.4.3　空间和灰度分辨率 …… 34
2.4.4　图像内插 …… 36
2.5　像素间的一些基本关系 …… 38
2.5.1　相邻像素 …… 38
2.5.2　邻接性、连通性、区域和边界 …… 38
2.5.3　距离度量 …… 40
2.6　数字图像处理中所用数学工具的介绍 …… 41
2.6.1　阵列与矩阵操作 …… 41
2.6.2　线性操作与非线性操作 …… 42
2.6.3　算术操作 …… 42
2.6.4　集合和逻辑操作 …… 46
2.6.5　空间操作 …… 49
2.6.6　向量与矩阵操作 …… 53
2.6.7　图像变换 …… 54
2.6.8　概率方法 …… 56
小结 …… 57
参考文献 …… 58
习题 …… 58
第3章　灰度变换与空间滤波 …… 62
引言 …… 62
3.1　背景知识 …… 62
3.1.1　灰度变换和空间滤波基础 …… 62
3.1.2　关于本章中的例子 …… 63
3.2　一些基本的灰度变换函数 …… 64
3.2.1　图像反转 …… 64
3.2.2　对数变换 …… 64
3.2.3　幂律(伽马)变换 …… 66
3.2.4　分段线性变换函数 …… 68
3.3　直方图处理 …… 72
3.3.1　直方图均衡 …… 72
3.3.2　直方图匹配(规定化) …… 77
3.3.3　局部直方图处理 …… 83
3.3.4　在图像增强中使用直方图统计 …… 85
3.4　空间滤波基础 …… 88
3.4.1　空间滤波机理 …… 88
3.4.2　空间相关与卷积 …… 89
3.4.3　线性滤波的向量表示 …… 92
3.4.4　空间滤波器模板的产生 …… 93
3.5　平滑空间滤波器 …… 93
3.5.1　平滑线性滤波器 …… 93
3.5.2　统计排序(非线性)滤波器 …… 96
3.6　锐化空间滤波器 …… 97
3.6.1　基础 …… 97
3.6.2　使用二阶微分锐化图像——拉普拉斯算子 …… 99
3.6.3　非锐化掩蔽和高提升滤波 …… 100
3.6.4　使用一阶微分锐化(非线性)图像——梯度 …… 101
3.7　混合空间增强法 …… 103
3.8　使用模糊技术进行灰度变换和空间滤波 …… 105
3.8.1　引言 …… 106
3.8.2　模糊集合论原理 …… 106
3.8.3　模糊集合应用 …… 110
3.8.4　使用模糊集合进行灰度变换 …… 116

3.8.5 使用模糊集合进行空间滤波……117
小结……119
参考文献……119
习题……120
第 4 章 频率域滤波……124
引言……124
4.1 背景……124
4.1.1 傅里叶级数和变换简史……124
4.1.2 关于本章中的例子……125
4.2 基本概念……125
4.2.1 复数……125
4.2.2 傅里叶级数……126
4.2.3 冲激及其取样特性……126
4.2.4 连续变量函数的傅里叶变换……128
4.2.5 卷积……130
4.3 取样和取样函数的傅里叶变换……131
4.3.1 取样……131
4.3.2 取样函数的傅里叶变换……132
4.3.3 取样定理……134
4.3.4 混淆……135
4.3.5 由取样后的数据重建(复原)函数……137
4.4 单变量的离散傅里叶变换(DFT)……138
4.4.1 由取样后的函数的连续变换得到 DFT……138
4.4.2 取样和频率间隔间的关系……140
4.5 两个变量的函数的扩展……141
4.5.1 二维冲激及其取样特性……141
4.5.2 二维连续傅里叶变换对……141
4.5.3 二维取样和二维取样定理……142
4.5.4 图像中的混淆……143
4.5.5 二维离散傅里叶变换及其反变换……147
4.6 二维离散傅里叶变换的一些性质……148
4.6.1 空间和频率间隔的关系……148
4.6.2 平移和旋转……148
4.6.3 周期性……148
4.6.4 对称性……150
4.6.5 傅里叶谱和相角……154
4.6.6 二维卷积定理……157
4.6.7 二维离散傅里叶变换性质的小结……159
4.7 频率域滤波基础……161
4.7.1 频率域的其他特性……161
4.7.2 频率域滤波基础……162
4.7.3 频率域滤波步骤小结……165
4.7.4 空间和频率域滤波间的对应……166
4.8 使用频率域滤波器平滑图像……169
4.8.1 理想低通滤波器……169
4.8.2 布特沃斯低通滤波器……172
4.8.3 高斯低通滤波器……173
4.8.4 低通滤波的其他例子……174
4.9 使用频率域滤波器锐化图像……176
4.9.1 理想高通滤波器……176
4.9.2 布特沃斯高通滤波器……178
4.9.3 高斯高通滤波器……178
4.9.4 频率域的拉普拉斯算子……179
4.9.5 钝化模板、高提升滤波和高频强调滤波……180
4.9.6 同态滤波……182
4.10 选择性滤波……184
4.10.1 带阻滤波器和带通滤波器……184
4.10.2 陷波滤波器……185
4.11 实现……187
4.11.1 二维 DFT 的可分性……187
4.11.2 用 DFT 算法计算 IDFT……187
4.11.3 快速傅里叶变换(FFT)……187
4.11.4 关于滤波器设计的一些注释……190
小结……190
参考文献……190
习题……191
第 5 章 图像复原与重建……196
引言……196
5.1 图像退化/复原过程的模型……197
5.2 噪声模型……197
5.2.1 噪声的空间和频率特性……197
5.2.2 一些重要的噪声概率密度函数……198
5.2.3 周期噪声……201
5.2.4 噪声参数的估计……202
5.3 只存在噪声的复原——空间滤波……203
5.3.1 均值滤波器……203
5.3.2 统计排序滤波器……205
5.3.3 自适应滤波器……208
5.4 用频率域滤波消除周期噪声……211
5.4.1 带阻滤波器……211
5.4.2 带通滤波器……211
5.4.3 陷波滤波器……212
5.4.4 最佳陷波滤波……213
5.5 线性、位置不变的退化……216
5.6 估计退化函数……218
5.6.1 图像观察估计……218
5.6.2 试验估计……218
5.6.3 建模估计……219
5.7 逆滤波……221
5.8 最小均方误差(维纳)滤波……222
5.9 约束最小二乘方滤波……224
5.10 几何均值滤波……227
5.11 由投影重建图像……228
5.11.1 引言……228

5.11.2 计算机断层(CT)原理 230
5.11.3 投影和雷登变换 232
5.11.4 傅里叶切片定理 235
5.11.5 使用平行射线束滤波反投影的重建 236
5.11.6 使用扇形射线束滤波反投影的重建 240
小结 244
参考文献 244
习题 245

第6章 彩色图像处理 249
引言 249
6.1 彩色基础 249
6.2 彩色模型 254
6.2.1 RGB 彩色模型 254
6.2.2 CMY 和 CMYK 彩色模型 257
6.2.3 HSI 彩色模型 257
6.3 伪彩色图像处理 262
6.3.1 灰度分层 262
6.3.2 灰度到彩色的变换 265
6.4 全彩色图像处理基础 267
6.5 彩色变换 268
6.5.1 公式 269
6.5.2 补色 271
6.5.3 彩色分层 271
6.5.4 色调和彩色校正 273
6.5.5 直方图处理 275
6.6 平滑和锐化 276
6.6.1 彩色图像平滑 276
6.6.2 彩色图像锐化 278
6.7 基于彩色的图像分割 279
6.7.1 HSI 彩色空间的分割 279
6.7.2 RGB 向量空间中的分割 279
6.7.3 彩色边缘检测 281
6.8 彩色图像中的噪声 283
6.9 彩色图像压缩 284
小结 285
参考文献 285
习题 286

第7章 小波和多分辨率处理 289
引言 289
7.1 背景 289
7.1.1 图像金字塔 290
7.1.2 子带编码 292
7.1.3 哈尔变换 297
7.2 多分辨率展开 300
7.2.1 级数展开 300
7.2.2 尺度函数 301
7.2.3 小波函数 304
7.3 一维小波变换 306
7.3.1 小波级数展开 306
7.3.2 离散小波变换 308
7.3.3 连续小波变换 309
7.4 快速小波变换 311
7.5 二维小波变换 317
7.6 小波包 322
小结 330
参考文献 330
习题 331

第8章 图像压缩 334
引言 334
8.1 基础知识 335
8.1.1 编码冗余 336
8.1.2 空间冗余和时间冗余 337
8.1.3 不相关的信息 337
8.1.4 图像信息的度量 338
8.1.5 保真度准则 340
8.1.6 图像压缩模型 341
8.1.7 图像格式、容器和压缩标准 343
8.2 一些基本的压缩方法 345
8.2.1 霍夫曼编码 345
8.2.2 Golomb 编码 346
8.2.3 算术编码 350
8.2.4 LZW 编码 351
8.2.5 行程编码 353
8.2.6 基于符号的编码 357
8.2.7 比特平面编码 359
8.2.8 块变换编码 361
8.2.9 预测编码 373
8.2.10 小波编码 387
8.3 数字图像水印 394
小结 398
参考文献 398
习题 399

第9章 形态学图像处理 402
引言 402
9.1 预备知识 402
9.2 腐蚀和膨胀 404
9.2.1 腐蚀 404
9.2.2 膨胀 406
9.2.3 对偶性 407
9.3 开操作与闭操作 407
9.4 击中或击不中变换 411
9.5 一些基本的形态学算法 412
9.5.1 边界提取 412
9.5.2 孔洞填充 413
9.5.3 连通分量的提取 414
9.5.4 凸壳 416

9.5.5 细化……417
9.5.6 粗化……418
9.5.7 骨架……418
9.5.8 裁剪……420
9.5.9 形态学重建……421
9.5.10 二值图像形态学操作小结……426
9.6 灰度级形态学……428
9.6.1 腐蚀和膨胀……428
9.6.2 开操作和闭操作……430
9.6.3 一些基本的灰度级形态学算法……431
9.6.4 灰度级形态学重建……435
小结……437
参考文献……437
习题……438

第 10 章 图像分割……443

引言……443
10.1 基础知识……443
10.2 点、线和边缘检测……445
10.2.1 背景知识……445
10.2.2 孤立点的检测……447
10.2.3 线检测……449
10.2.4 边缘模型……450
10.2.5 基本边缘检测……454
10.2.6 更先进的边缘检测技术……459
10.2.7 边缘连接和边界检测……467
10.3 阈值处理……476
10.3.1 基础知识……476
10.3.2 基本的全局阈值处理……478
10.3.3 用 Otsu 方法的最佳全局阈值处理……479
10.3.4 用图像平滑改善全局阈值处理……483
10.3.5 利用边缘改进全局阈值处理……484
10.3.6 多阈值处理……487
10.3.7 可变阈值处理……488
10.3.8 多变量阈值处理……492
10.4 基于区域的分割……493
10.4.1 区域生长……493
10.4.2 区域分裂与聚合……495
10.5 用形态学分水岭的分割……497
10.5.1 背景知识……497
10.5.2 水坝构建……498
10.5.3 分水岭分割算法……499
10.5.4 标记的使用……501
10.6 分割中运动的应用……502
10.6.1 空间域技术……502
10.6.2 频率域技术……504
小结……507
参考文献……507
习题……508

第 11 章 表示和描述……514

引言……514
11.1 表示……514
11.1.1 边界追踪……514
11.1.2 链码……516
11.1.3 使用最小周长多边形的多边形近似……518
11.1.4 其他多边形近似方法……522
11.1.5 标记图……523
11.1.6 边界线段……525
11.1.7 骨架……525
11.2 边界描绘子……527
11.2.1 一些简单的描绘子……527
11.2.2 形状数……528
11.2.3 傅里叶描绘子……529
11.2.4 统计矩……531
11.3 区域描绘子……532
11.3.1 一些简单的描绘子……532
11.3.2 拓扑描绘子……532
11.3.3 纹理……534
11.3.4 不变矩……542
11.4 使用主成分进行描绘……544
11.5 关系描绘子……550
小结……553
参考文献……553
习题……554

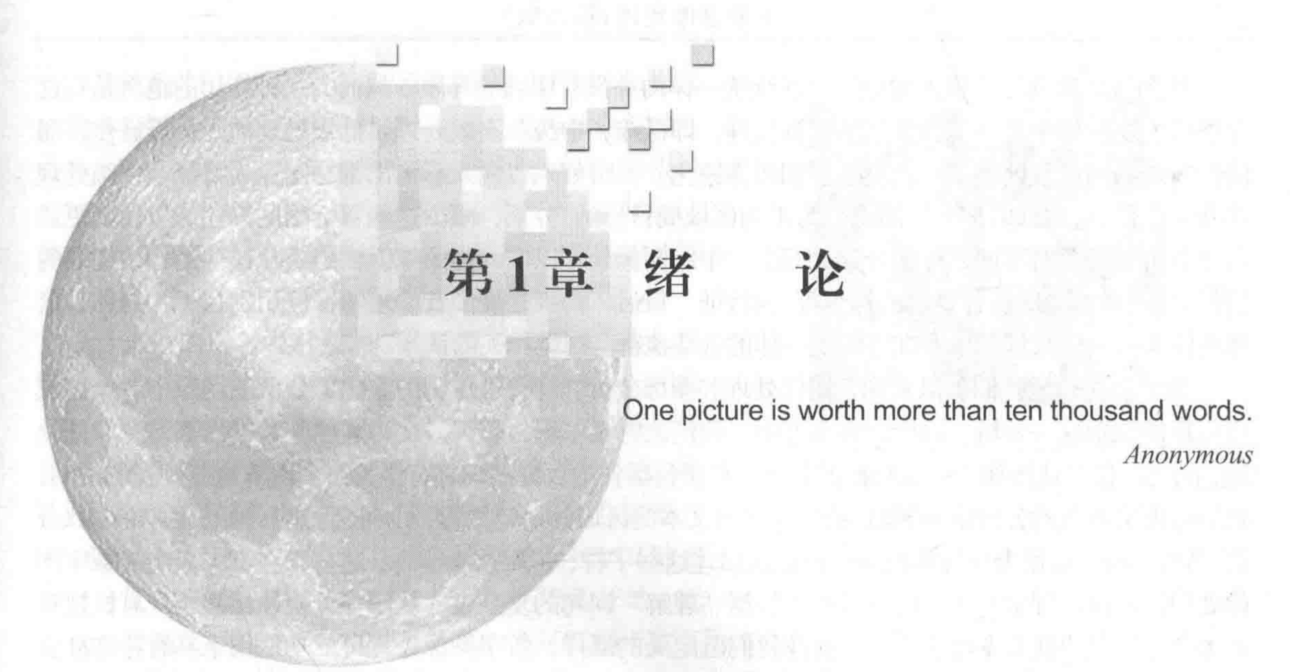
第 12 章 目标识别……557

引言……557
12.1 模式和模式类……557
12.2 基于决策理论方法的识别……560
12.2.1 匹配……560
12.2.2 最佳统计分类器……564
12.2.3 神经网络……570
12.3 结构方法……585
12.3.1 匹配形状数……585
12.3.2 串匹配……586
小结……587
参考文献……588
习题……588

附录 A 图像压缩编码表……591

附录 B 参考书目……595

索引……620



第1章 绪 论

One picture is worth more than ten thousand words.

Anonymous

引言

数字图像处理方法的重要性源于两个主要应用领域：改善图示信息以便人们解释；为存储、传输和表示而对图像数据进行处理，以便于机器自动理解。本章有几个主要目的：(1)定义我们称之为图像处理领域的范围；(2)从历史观点回顾图像处理的起源；(3)考察一些主要的应用领域，给出图像处理技术状况的概念；(4)简要讨论数字图像处理中所用的主要方法；(5)概述典型通用图像处理系统的组成；(6)列出数字图像处理领域已公开发表的一些图书和文献。

1.1 什么是数字图像处理

一幅图像可定义为一个二维函数 $f(x, y)$，其中 x 和 y 是空间(平面)坐标，而在任何一对空间坐标 (x, y) 处的幅值 f 称为图像在该点处的强度或灰度。当 x, y 和灰度值 f 是有限的离散数值时，我们称该图像为数字图像。数字图像处理是指借助于数字计算机来处理数字图像。注意，数字图像是由有限数量的元素组成的，每个元素都有一个特定的位置和幅值。这些元素称为图画元素、图像元素或像素。像素是广泛用于表示数字图像元素的术语。在第2章，我们将用更正式的术语来考虑这些定义。

视觉是人类最高级别的感知，所以图像在人类感知中扮演着最重要的角色。然而，人类的感知仅限于电磁波谱的可见光波段；与人类不同，成像机器几乎可以覆盖从伽马射线到无线电波的整个电磁波谱范围。它们可以对人类不习惯的那些图像源进行加工，包括超声波、电子显微镜和计算机产生的图像。因而，数字图像处理涉及很宽泛的各种应用领域。

关于图像处理止于何处或其他相关领域（如图像分析和计算机视觉)从何处开始，创始人之间并没有一致的看法。有时，用输入和输出内容都是图像这一规范来界定图像处理的范围。我们认为这有点局限，是人为的界定。例如，在这种定义下，甚至计算一幅图像的平均灰度(结果为一个数字)这样简单的任务都不能算是图像处理。另一方面，有些领域(如计算机视觉)的最终目标是使用计算机来模拟人的视觉，包括理解并根据视觉输入采取动作等。该领域本身是人工智能(AI)的一个分支，其目的是模仿人类智能。人工智能领域正处于发展过程的初级阶段，其发展要比预期的慢得多。图像分析(也称为图像理解)领域则处在图像处理和计算机视觉之间。

① 每页下方阴影框中的数字表示译文对应的原书页码，索引中的页码为原书页码。——编者注

从图像处理到计算机视觉的这个连续统一体内并没有明确的界限。然而，一种有用的范例是在这个连续的统一体中考虑三种典型的计算处理，即低级、中级和高级处理。低级处理涉及初级操作，如降低噪声的图像预处理、对比度增强和图像锐化。低级处理以输入、输出都是图像为特征。中级处理涉及诸多任务，譬如(把一幅图像分为不同区域或目标的)分割，减少这些目标物的描述，以使其更适合计算机处理及对不同目标的分类(识别)。中级图像处理以输入为图像但输出是从这些图像中提取的特征(如边缘、轮廓及各物体的标识等)为特征。最后，高级处理涉及“理解”已识别目标的总体，就像在图像分析中那样，以及在连续统一体的远端执行与视觉相关的认知功能。

基于上述讨论我们可以看到，图像处理和图像分析之间合乎逻辑的重叠部分就是图像中各个区域或目标的识别这一领域。这样，在本书中，我们就将数字图像处理界定为其输入和输出都是图像的处理，另外，包含从图像中提取特征的处理，直至包括各个目标的识别。作为一个澄清这些概念的说明，我们考虑文本自动分析这一领域。获取包含有文本的区域的一幅图像，对该图像进行预处理，以提取(分割)各个字符，以适合于计算机处理的形式描述这些字符，识别这些字符，这些都在本书界定的数字图像处理范畴内。理解一页的内容可能要根据“理解”词句的复杂度，从图像分析领域甚至计算机视觉来考虑。这很快就会变得很明显，就像我们已定义的那样，数字图像处理可成功地用于具有特殊社会和经济价值的广泛领域。后面各章中给出的概念是这些应用领域中所用方法的基础。

1.2 数字图像处理的起源

数字图像的最早应用之一是在报纸业，当时，图片第一次通过海底电缆从伦敦传往纽约。早在20世纪20年代曾引入巴特兰(Bartlane)电缆图片传输系统，把横跨大西洋传送一幅图片所需的时间从一个多星期减少到了3小时。为了用电缆传输图片，首先使用特殊的打印设备对图片编码，然后在接收端重构这些图片。图1.1就是用这种方法传送并利用装有打印机字体的电报打印机模拟中间色调还原的图像。

这些早期数字图片视觉质量的改进中的初始问题涉及打印过程的选择和亮度等级的分布。用于得到图1.1所示图像的打印方法到1921年底就被彻底淘汰，转而支持一种基于照相还原的技术，即在电报接收端使用穿孔纸带来还原图片。图1.2显示了使用这种方法得到的一幅图像。与图1.1相比，它在色调质量和分辨率方面的改进都很明显。

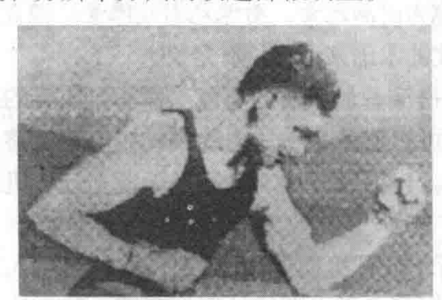


图1.1 1921年由电报打印机采用特殊字体在编码纸带上产生的数字图片(原图像由McFarlane提供①)

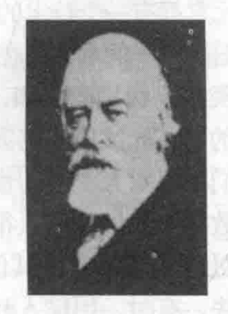


图1.2 1922年在信号两次穿越大西洋后，由穿孔纸带得到的数字图片(原图像由McFarlane提供)

早期的巴特兰系统可以使用5个不同的灰度级来编码图像。到1929年，这一能力已增大到15级。图1.3所示的这种典型图像就是用15级色调设备得到的。在这一时期，由于引入了一种使用编码图片纸带调制光束而使底片感光的系统，因而明显地改善了还原过程。

24~26

① 本书后面提供的参考文献按作者姓氏的字母顺序列出。

虽然刚才引用的例子中涉及数字图像，但并不认为它们就是我们定义的数字图像处理，因为创建这些图像时并未涉及计算。因此，数字图像处理的历史与数字计算机的发展密切相关。事实上，数字图像要求非常大的存储和计算能力，因此数字图像处理领域的发展必须依靠数字计算机及数据存储、显示和传输等相关支撑技术的发展。

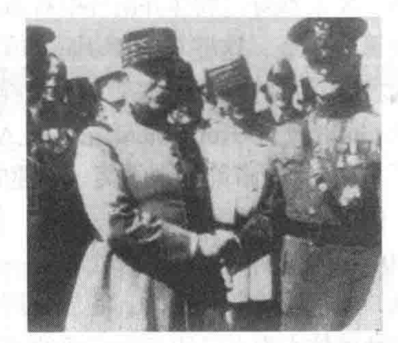


图 1.3　美国将军潘兴和法国元帅福熙的照片，该照片未加修饰，是在 1929 年从伦敦到纽约使用 15 级色调设备通过电缆传送的（原图像由 McFarlane 提供）

计算机的概念可追溯到 5000 多年前亚洲算盘的发明。更近一些，过去两个世纪以来已有的发展也奠定了我们今天称之为计算机的基础。然而，我们称之为现代计算机的基础还要回溯到 20 世纪 40 年代由约翰·冯·诺依曼提出的两个重要概念：(1) 保存程序和数据的存储器；(2) 条件分支。这两个概念是中央处理单元（CPU）的基础，今天，它是计算机的心脏。从冯·诺依曼开始，已有的一系列重要的进展使得计算机强大到足以用于数字图像处理。

简单地说，这些进展可归纳为如下几点：(1) 1948 年美国贝尔实验室发明了晶体管；(2) 20 世纪 50 年代和 60 年代高级编程语言 COBOL（Common Business-Oriented Language）及 FORTRAN（Formula Translator）的开发；(3) 1958 年美国德州仪器公司发明了集成电路（IC）；(4) 20 世纪 60 年代早期操作系统的开发；(5) 20 世纪 70 年代早期 Intel 公司开发了微处理器（由中央处理单元、存储器和输入/输出控制组成的单一芯片）；(6) 1981 年 IBM 公司推出了个人计算机；(7) 元器件的逐步小型化，随着 20 世纪 70 年代末大规模集成电路（LI）开始出现，20 世纪 80 年代出现了甚大规模集成电路（VLSI），今天出现了超大规模集成电路（ULSI）。伴随着这些进展，数字图像处理的两个基本需求——大容量存储和显示系统领域也随之快速发展。

第一台功能强大到足以执行有意义图像处理任务的大型计算机出现在20世纪60年代初。我们今天称之为数字图像处理的诞生可追溯至这一时期这些机器的使用和空间项目的开发。这两大发展一起把人们的注意力集中到数字图像处理概念的潜能上。利用计算机技术改善空间探测器发回的图像的工作，始于 1964 年美国加利福尼亚的喷气推进实验室。当时由“徘徊者 7 号”卫星传送的月球图像由一台计算机进行了处理，以校正航天器上电视摄像机中各种类型的图像畸变。图 1.4 显示了由“徘徊者 7 号”于 1964 年 7 月 31 日上午（东部白天时间）9 点 09 分在撞击月球表面前约 17 分钟时拍摄的第一张月球图像［网状标记用于第 2 章讨论的几何校正］。这也是由美国航天器拍摄的第一幅月球图像。“徘徊者 7 号”积累的图像处理方法可作为增强和复原图像方法的基础，譬如“探索者”登月飞行、“水手号”火星飞越任务及“阿波罗”载人登月和其他任务。

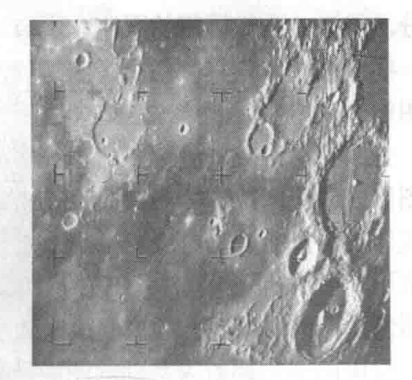


图 1.4　美国航天器传回的第一张月球照片，它是由“徘徊者 7 号”探测器在撞击月球表面前 17 分钟于美国东部时间 1964 年 7 月 31 日上午 9 点 09 分拍摄的（原图像由 NASA 提供）

在空间应用的同时，数字图像处理技术在 20 世纪 60 年代末和 20 世纪 70 年代初开始用于医学成像、地球资源遥感监测和天文学等领域。早在 20 世纪 70 年代发明的计算机轴向断层术，简称为计算机断层（CT），是图像处理在医学诊断领域最重要的应用之一。计算机轴向断层术是一种处理方法，在这种处理中，检测器环围绕一个物体（或病人），一个与该环同心的 X 射线源（与检测器环同心）绕物

26～28

体旋转。X射线穿过物体并由环中对面的检测器进行收集。当X射线源旋转时，重复这一过程。断层由一些算法组成，这些算法使用感知的数据来重建通过物体的“切片”图像。当物体沿垂直于检测器环的方向运动时，就产生一系列这样的“切片”，这些切片组成该物体内部的三维再现。断层摄影术是由Godfrey N. Hounsfield先生和Allan M. Cormack教授分别发明的。由于这项发明，他们共同获得了1979年的诺贝尔医学奖。有趣的是，X射线是1895年由Wilhelm Conrad Roentgen(威廉·康拉德·伦琴)发现的，由于这一发现，他获得了1901年的诺贝尔物理学奖。今天，这两个时间上相差近100年的发明引领着图像处理的一些最重要应用。

从20世纪60年代至今，图像处理领域一直在生机勃勃地发展。除了医学和空间项目应用外，数字图像处理技术现在已用于更广泛的范围。计算机方法用于增强对比度或将灰度编码为彩色，以便于解释工业、医学及生物科学等领域中的X射线图像和其他图像。地理学者使用相同或相似的技术，由航空和卫星图像研究污染模式。图像增强和复原方法用于处理不可修复物体的退化图像，或太昂贵以至于不可复制的实验结果。在考古学领域，使用图像处理方法已成功地复原了模糊的图片，这些图片是丢失或损坏的稀有物品的唯一现有记录。在物理学和相关领域，计算机技术通常用于增强如高能等离子和电子显微镜等领域的实验图像。类似地，图像处理技术也成功地应用在天文学、生物学、核医学、法律实施、国防和工业领域中。

这些例子说明图像处理的结果主要用于人类解译。本章开始时曾提到，数字图像处理技术的第二个主要应用领域是解决机器感知的问题。在这种情况下，兴趣在于以更适合计算机处理的形式从图像中提取信息的过程。通常，这种信息类似于人类用于解译一幅图像内容的视觉特性。例如，机器感知中使用的信息类型的例子有统计矩、傅里叶变换系数和多维距离度量。在机器感知中，使用图像处理技术的典型问题是自动字符识别、产品装配线和检测的工业机器视觉、军事识别、指纹的自动处理、X射线和血样分类、用于天气预报和环境评估的航空图像与卫星图像的机器处理。计算机价格性能比的不断下降、万维网和互联网规模的不断扩张及网络通信带宽的提高，为数字图像处理技术的持续发展提供了前所未有的机会。下一节中将说明某些这样的应用领域。

1.3　使用数字图像处理领域的实例

今天，几乎不存在与数字图像处理无关的技术领域。这里讨论的范围只能覆盖其应用领域的一小部分。然而，由于篇幅的限制，本节的内容无疑将围绕数字图像处理的广度和重要性展开。本节将介绍一些应用领域，其中每个领域都要使用后续各章中讲解的常用数字图像处理技术。本节中所示的许多图像会在本书后面给出的多个例子中引用。显示的所有图像都是数字图像。

数字图像处理的应用领域多种多样，所以本书在组织形式上力图覆盖该技术领域的广度。阐述数字图像处理应用范围的一种最简方法是，根据信息源来分类(如可见光或X射线等)。在今天的应用中，最主要的图像能源是电磁能谱，其他主要图像能源包括声波、超声波和电子(以用于电子显微镜中的电子束形式)。用于建模和可视化的合成图像由计算机产生。本节简要讨论如何生成各种类别的图像及这些图像适用的领域。把图像变换为数字图像的方法则在下一章中讨论。

以电磁波谱辐射为基础的图像是我们最熟悉的，特别是X射线和可见光谱波段的图像。电磁波可定义为以各种波长传播的正弦波，或视为无质量的粒子流，每个粒子以波的形式传播并以光的速度运动。每个无质量的粒子包含一定的能量(或一束能量)，每束能量称为一个光子。如果光谱波段根据光子能量进行分组，则可得到图1.5所示的光谱，范围从伽马射线(最高能量)到无线电波(最低能量)。如图所示，加底纹的条带表明了这样一个事实，即电磁波谱的各个波段之间并没有明确的界线，而是由一个波段平滑地过渡到另一个波段。

28~29

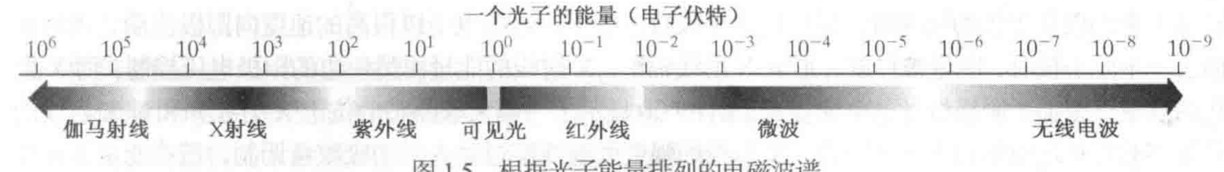


图 1.5　根据光子能量排列的电磁波谱

1.3.1　伽马射线成像

伽马射线成像的主要用途包括核医学和天文观测。在核医学中，这种方法是将放射性同位素注射到人体内，当这种物质衰变时就会放射出伽马射线，然后用伽马射线检测仪收集到的放射线来产生图像。图1.6(a)显示了一幅使用伽马射线成像得到的人体骨骼扫描图像。这类图像用于骨骼病变的定位，譬如感染或肿瘤。图 1.6(b)显示了另一种主要形态的核成像，称为*正电子放射断层*(PET)，其原理与1.2 节中提及的 X 射线断层相同。然而，与使用外部 X 射线源不同，它给病人注入放射性同位素，同位素衰变时放射出正电子。正电子遇到一个电子时，两者湮没并放射出两束伽马射线。检测到这些射线后，就可利用断层技术的基本原理创建断层图像。图 1.6(b)所示的图像是构成病人三维再现图像序列中的一幅图像。这幅图像表明脑部和肺部各有一个肿瘤，即很容易看到的白色小团块。

大约在15 000年前，天鹅星座中的星球发生大爆炸，产生了一团过热的稳定气云(即天鹅星座环)，该气云以彩色阵列形式发光。图1.6(c)显示了在伽马射线波段成像的天鹅星座环。与图1.6(a)和图1.6(b)不同，该图像是利用被成像物体的自然辐射得到的。最后，图 1.6(d)显示了一幅来自核反应堆真空管的伽马辐射图像，在图像的左下方可以看到较强的辐射区。

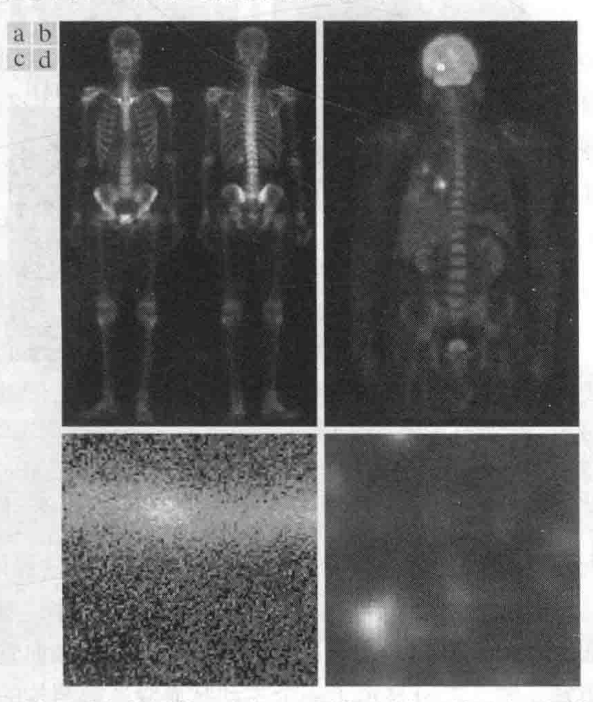


图 1.6　伽马射线成像实例：(a)骨骼扫描图像；(b)PET 图像；(c)天鹅星座环图像；(d)来自反应堆真空管的伽马辐射(亮点)［图(a)由 G. E. Medical Systems 公司提供，图(b)由 CTI PET Systems 公司的 Michael E. Casey 博士提供，图(c)由 NASA 提供，图(d)由密歇根大学的 Zhong He 和 David K.Wehe 教授提供］

1.3.2　X 射线成像

X 射线是最早用于成像的电磁辐射源之一。最熟悉的 X 射线应用是医学诊断，但是 X 射线还被广泛用于工业和其他领域，如天文学。用于医学和工业成像的 X 射线是由 X 射线管产生的，X 射线

30～31

管是带有阴极和阳极的真空管。阴极加热释放自由电子，这些电子以很高的速度向阳极流动，当电子撞击一个原子核时，能量被释放并形成 X 射线辐射。X 射线的能量由另一边的阳极电压控制，而 X 射线的数量由施加于阴极灯丝的电流控制。图1.7(a)显示了一幅大家熟悉的位于 X 射线源和对 X 射线能量敏感的胶片之间的病人胸部图像。X 射线的强度受射线穿过病人时的吸收量调制，最终能量落在胶片上并使胶片感光，这与光使得胶片感光的原理是一样的。在数字射线照相术中，数字图像可用两种方法得到：(1)使用数字化的 X 射线胶片；(2)X 射线穿过病人身体后直接落到某个装置上(譬如荧光屏)，该装置把 X 射线转换为光信号。然后，转换而来的光信号由高灵敏度的数字系统捕获。第 2 章和第 4 章将详细地讨论数字化问题。

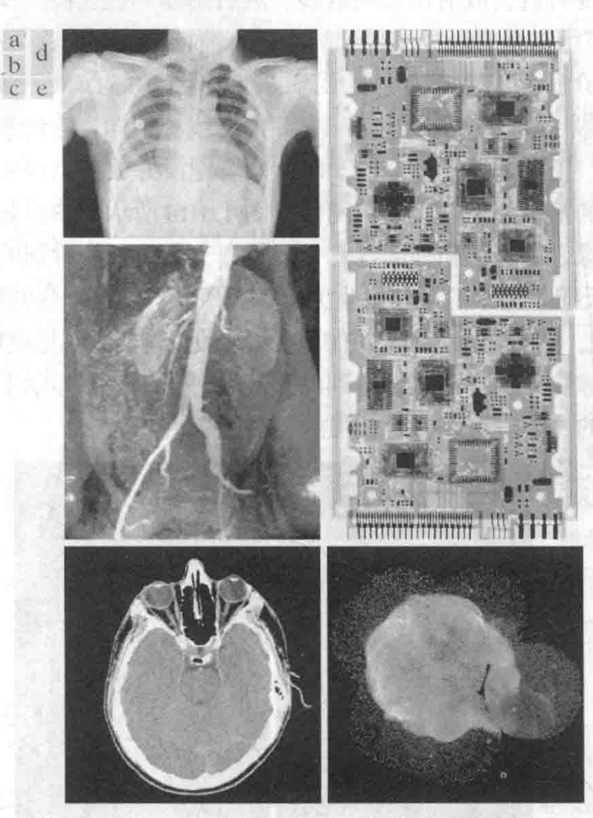


图 1.7　X 射线成像实例：(a)胸部 X 射线图像；(b)主动脉造影图像；(c)头部 CT 图像；(d)电路板图像；(e)天鹅星座环图像［图(a)和图(c)由 Vanderbilt 大学医学中心辐射学与放射学系的 David R. Pickens 博士提供，图(b)由密歇根大学医学院解剖学分部的 Thomas R. Gest 博士提供，图(d)由 Lixi 公司的 Joseph E. Pascente 先生提供，图(e)由 NASA 提供］

血管照相术是对比度增强辐射成像领域中的另一个主要应用。该过程用于得到血管的图像(称为血管造影照片)。一根导管(柔软且中空的小管)插入动脉或静脉，导管穿过血管并被引导到要研究的区域。当导管到达所研究的部位时，将 X 射线造影剂注入导管。这会增强血管的对比度，并可以让放射线学者观察到任何病变或阻塞。图 1.7(b)显示了一个主动脉血管造影照片的例子。在图像的左下方，可以看到插入到血管中的导管。注意，在图像中，可看到造影剂流向肾脏时大血管的强烈对比效果。正如第 2 章将要讨论的那样，血管照相术是数字图像处理的主要应用领域，在该领域，图像相减用于进一步增强被研究的血管。

X 射线在医学成像中的另一个重要应用是计算机轴向断层(CAT)。由于该技术的分辨率和三维能力，CAT 扫描早在 20 世纪 70 年代第一次付诸使用时就引起了医疗手段的革命。正如 1.2 节提到的那样，每幅 CAT 图像都是垂直穿过病人的一个“切片”，当病人纵向移动时可产生大量的“切片”，这些

31
~
33

图像组合在一起就构成了人体内部的三维描绘图像，其纵向分辨率与切片数量成正比。图 1.7(c) 显示了一幅典型的头部 CAT 切片图像。

与刚刚讨论的有些类似的技术也可用于工业处理，但通常使用的是更高能量的 X 射线。图1.7(d) 显示了一块电路板的 X 射线图像。这样的图像是 X 射线在上百种典型工业应用中的代表性图像，用于检测电路板中的制造缺陷，如元件缺失或断线等。当元件可被 X 射线穿透时，CAT 扫描很有用，譬如塑料元件，甚至更大的物体，如固体推进剂火箭发动机。图1.7(e) 显示了天文学中 X 射线成像的一个例子。这幅图像是图1.6(c) 中的天鹅星座环，但此时是 X 射线波段成像。

1.3.3　紫外波段成像

紫外“光”的应用多种多样，包括平板印刷术、工业检测、显微方法、激光、生物成像和天文观测等。我们仅用显微方法和天文观测作为例子来说明这一波段的成像。

紫外光用于荧光显微方法中，这是显微方法中发展最快的领域之一。荧光是在 19 世纪中叶发现的一种现象。当时，当紫外光直接照射到矿物质上时，首次发现了萤石发出的荧光。紫外光本身并不可见，但当紫外辐射光子与荧光材料内原子中的电子碰撞时，它把电子提高到较高的能级，随后受激电子释放到较低的能级，并以可见光范围内的低能光子形式发光。荧光显微方法的基本任务是用激发光照射一个样品，然后从较强的激发光中分离出较弱的荧光，这样，仅有辐射光到达人眼或其他检测器，以允许检测足够的对比度而得到照射在暗背景上的荧光区。非荧光材料的背景越暗，设备越有效。

荧光显微方法在研究可产生荧光的材料时，的确是一种很优秀的方法，无论是以它们的自然形式（原发荧光），还是经化学处理后具有荧光能力（次生荧光）。图1.8(a) 和图1.8(b) 显示了使用荧光显微方法所得到的典型结果。图1.8(a) 显示了普通玉米的荧光显微图像，图1.8(b) 显示了被一种谷类疾病“黑穗病”感染了的玉米图像，黑穗病是草、洋葱、高粱等被 700 多种寄生真菌感染后出现的病害。玉米黑穗病是一种严重的病害，因为玉米是世界上最主要的食品来源之一。图 1.8(c) 显示了在紫外波段的高能区域成像的天鹅星座环。

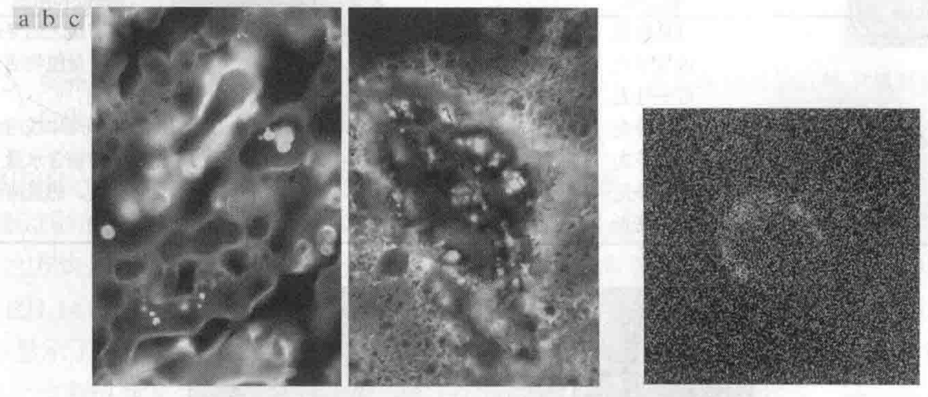


图 1.8　紫外光成像实例：(a) 普通玉米图像；(b) 患黑穗病的玉米图像；(c) 天鹅星座环图像
[图(a) 和图(b) 由佛罗里达州立大学的 Michael W. Davidson 博士提供，图(c) 由 NASA 提供]

1.3.4　可见光及红外波段成像

由于电磁波谱可见光波段在所有波段中是我们最熟悉的，因此这一波段的成像应用领域远远超过其他波段的成像应用领域。红外波段常用于与可见光相结合成像，为便于说明，本节把可见光和红外光合并在一起讨论。下面的讨论将涉及光显微方法、天文学、遥感、工业和法律实施等方面的应用。

33
~
34

图1.9显示了一些用可见光显微技术得到的图像。这些例子涵盖了从制药和微检测到材料特征的范围。甚至仅在显微方法中，其应用领域如此之多，以至于不能在这里详细列出。当然，把这些图像用于从增强到测量这些范围可能并不困难。

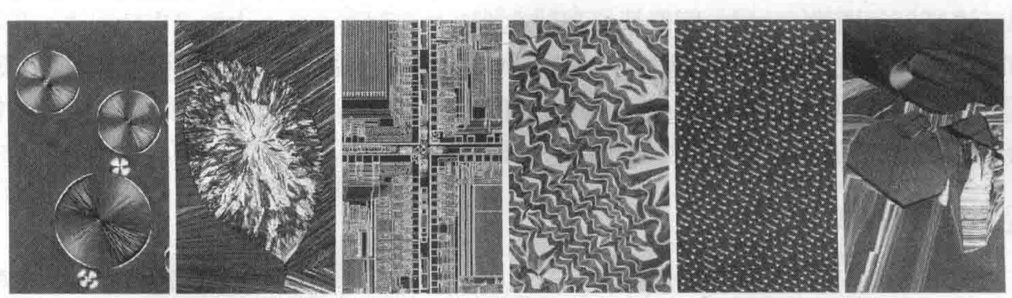


图 1.9　光显微镜图像实例：(a)已放大 250 倍的紫杉酚(抗癌剂)；(b)已放大 40 倍的胆固醇；(c)已放大 60 倍的微处理器；(d)已放大 600 倍的镍氧化物胶片；(e)已放大 1750 倍的音频 CD 表面；(f)已放大 450 倍的有机超导体(这些图像由佛罗里达州立大学的 Michael W. Davidson 博士提供)

可见光处理的另一个主要应用领域是遥感，遥感通常包括可见光和红外波谱范围的一些波段。表 1.1 显示了 NASA 的 LANDSAT 卫星中的主要波段。LANDSAT 卫星的主要功能是从空间获得并传送地球的图像，目的是监测地球的环境条件。波段用波长来表示，1 μm 等于 10^{-6} m(我们讨论的是电磁波谱的波长范围，第 2 章将详细讨论)。注意表 1.1 中每个波段的特征与用途。

为了阐述这类多光谱成像的基本性能，考虑图1.10，它为表1.1中的每个频谱波段显示了一幅图像。成像区域是美国华盛顿特区，包括建筑物、道路、植被和穿过城市的主要河流(波拖马可河)等特征。人口中心地区的图像常被用来评估人口增长和变迁方式、污染及其他有害环境的因素。可见光和红外光特征之间的区别在这些图像中十分明显。例如，在波段 4 和波段 5 中，从其周围环境可以看出如何较好地界定河流。

表 1.1　NASA 的 LANDSAT 卫星的主要波段

波段号	名　称	波长(μm)	特征和用途
1	可见蓝光	0.45~0.52	对水体有最大的穿透性
2	可见绿光	0.52~0.60	适用于度量植物活力
3	可见红光	0.63~0.69	植被辨别
4	近红外光	0.76~0.90	生物团和海岸线测绘
5	中红外光	1.55~1.75	土壤和植被含水量
6	热红外光	10.4~12.5	土壤温度，热量测绘
7	中红外光	2.08~2.35	矿物测绘

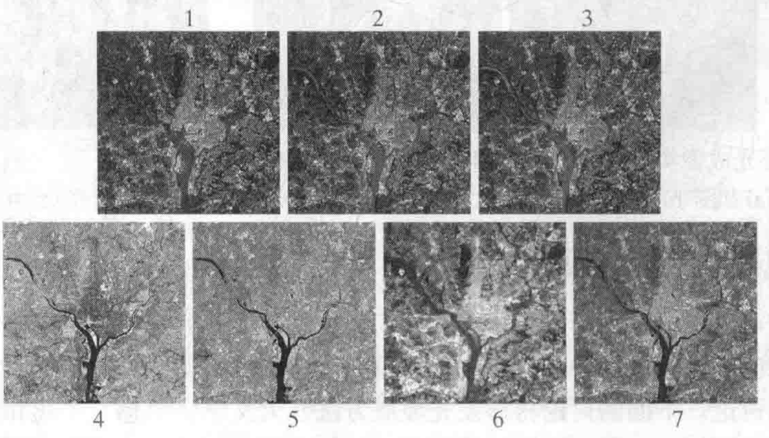


图 1.10　美国华盛顿特区的LANDSAT卫星图像。图中的数字与表 1.1 中的波段号对应(图像由 NASA 提供)

天气观测与预报也是卫星多光谱成像的主要应用领域。例如，图1.11是由美国国家海洋和大气管理(NOAA)卫星在可见光及红外波段使用传感器拍摄的飓风图像，图像中的飓风眼清晰可见。



图 1.11 2005年8月29日拍摄的“卡特里娜”飓风的卫星图像(图像由NOAA提供)

图1.12和图1.13显示了红外成像的一个应用。这些图像是“全球夜间灯光”数据集的一部分，它可以提供全球人类居住区的汇总情况。图像是由安装在NOAA DMSP(防卫气象卫星计划)卫星上的红外成像系统生成的。红外成像系统工作于10.0~13.4 μm波段，具有独特的功能，可观察地球表面微弱的接近可见光的近红外发射光，包括城市、小镇、村庄、气体火焰及火光。即使在图像处理方面没有正式地训练过，编写一个计算机程序，用这些图像来估计世界不同地区使用电能的百分比也并不困难。

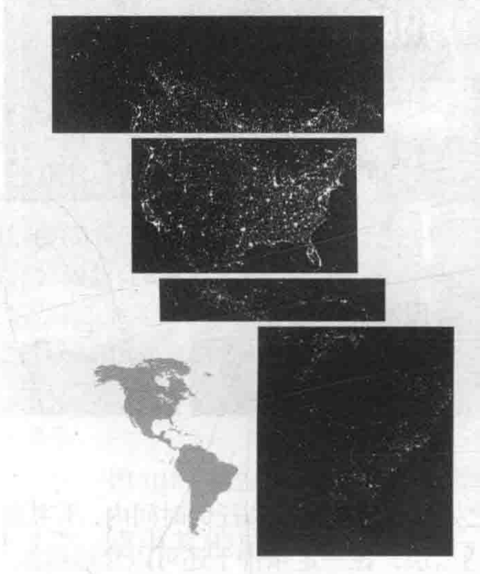


图 1.12 美洲的红外卫星图像。小灰色图仅供参考(图像由NOAA提供)

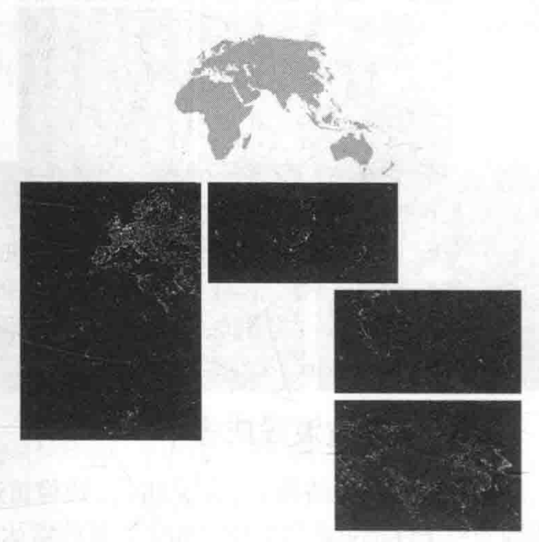


图 1.13 世界居住区的红外卫星图像。小灰色图仅供参考(图像由NOAA提供)

可见光谱中的一个主要成像领域是生产产品的自动视觉检测。图1.14显示了一些例子。图1.14(a)是CD-ROM驱动器的一个控制板。对这类产品的一种典型图像处理任务是检测丢失的部件(图像右上角1/4处的黑色方块是一个丢失部件的例子)。图1.14(b)是药丸胶囊的图像，这里的目标是用机器寻找缺失的药丸。图1.14(c)显示了一种应用，在该应用中，图像处理技术用来寻找未装满到要求液位的瓶子。图1.14(d)显示了一个清澈的塑料部件，部件内带有不可接受数量的气泡。这些异常检测是工业检测的主旨，这类检测也涉及其他产品，如木材和布匹。图1.14(e)显示了一批谷物，通过颜色和存在的异常(如变黑的小片)等进行检测。最后，图1.14(f)显示了一幅含有掺杂物的目镜图像(位于人眼处的镜头)，使用了一种“结构光”照明技术加以突出，以便更容易检测平面镜头中心处出现的镜头畸变。在类似钟表表盘的1点和5点处的刻痕是由钳子造成的损坏，而其他的多数小斑点则是残片。这类检测的目标是在包装之前自动找出损坏的或未正确加工的装配。

作为可见光谱中图像处理的最后说明，考虑图1.15。图1.15(a)显示了一幅拇指指纹图像。指纹图像通常由计算机处理，该处理不是增强指纹图像就是寻找特征，以便自动搜索数据库，寻找潜在的匹配。图1.15(b)显示了一幅纸币图像。数字图像处理在该领域的应用包括自动计数、法律实施、为

37~39

追踪和鉴别钱币读取序列号等。图 1.15(c)和图 1.15(d)所示的两幅汽车图像是自动读取车牌的例子。明亮的矩形指示成像系统检测车牌的区域。黑色矩形显示了系统自动读取车牌内容后的结果。车牌和其他字符识别广泛用于交通监测和监视中。

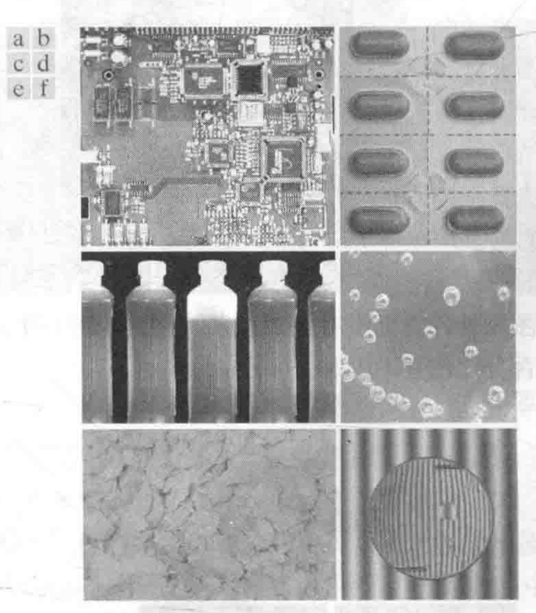


图 1.14 使用数字图像处理技术检查产品的一些例子：(a)一块电路板控制器图像；(b)封装的药丸图像；(c)瓶子图像；(d)清澈的塑料产品中的气泡图像；(e)谷物图像；(f)目镜搀杂物图像［图(f)由 Perceptics 公司的 Pete Sites 先生提供］

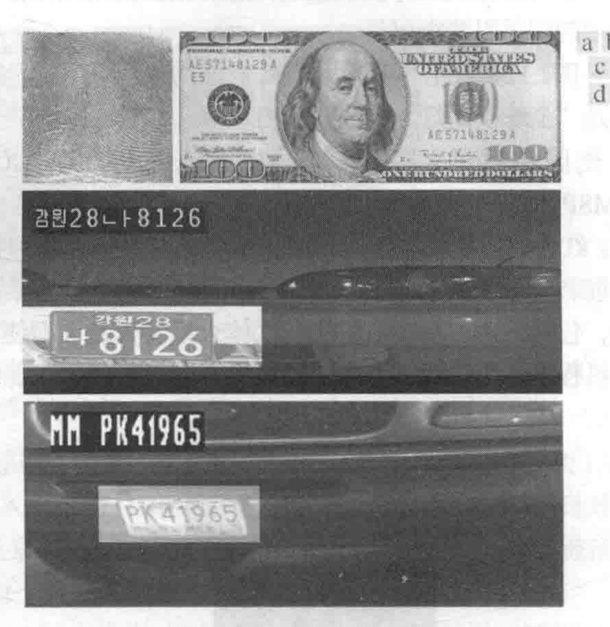


图 1.15 可见光谱成像的其他几个例子：(a)拇指指纹图像；(b)纸币图像；(c)~(d)自动读取车牌图像［图(a)由美国国家标准与技术研究所提供，图(c)和图(d)由 Perceptics 公司的 Juan Herrera 博士提供］

1.3.5 微波波段成像

微波波段成像的典型应用是雷达。成像雷达的独特之处是在任何范围和任何时间内，不考虑气候、周围光照条件都可收集数据的能力。某些雷达波可以穿透云层，在一定条件下还可以穿透植被、冰层和极干燥的沙漠。在许多情况下，雷达是探测地球表面不可接近地区的唯一方法。成像雷达的工作原理就像一台闪光照相机，它自己提供照明(微波脉冲)去照亮地面上的一个区域，并得到一幅快照图像。与照相机镜头不同，雷达使用天线和数字计算机记录图像。在雷达图像中，能看到的只是反射到雷达天线的微波能量。

图 1.16 显示了一幅西藏东南方高低不平山区的星载雷达图像，拉萨市约在东边 90 km 处，右下角是广阔的拉萨河谷，这是西藏农民和牧民包括曼巴居住的地方。这一地区的山峰海拔高度约为 5800 m，其谷底位于海拔 4300 m 处。注意，图像的清晰度和细节未被云层及其他大气条件阻挡，通常这些因素在可见光波段会影响图像。

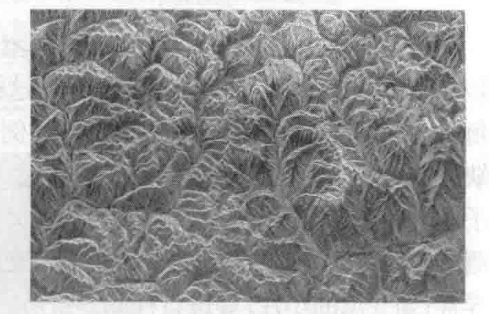


图 1.16 西藏东南部山区的星载雷达图像(图像由 NASA 提供)

1.3.6 无线电波段成像

40 ~ 42

正像波谱另一端(伽马射线)的成像情况那样，无线电波段成像主要应用于医学和天文学。在医学中，无线电波用于核磁共振成像(MRI)。该技术是把病人放在强磁场中，并让无线电波短脉冲通过

病人的身体，每个脉冲将导致由病人的组织发射的无线电响应脉冲，这些信号发生的位置和强度由计算机确定，从而产生病人的一幅二维剖面图像。MRI 可以在任何平面产生图像。图1.17 显示了人的膝盖和脊椎的图像。




图 1.17　人的 MRI 图像：(a) 膝盖图像；(b) 脊椎图像［图 (a) 由密歇根大学医学院解剖学分部的 Thomas R. Gest 博士提供，图 (b) 由 Vanderbilt 大学医学中心辐射学与放射学系的 David R. Pickens 博士提供］

图1.18 最右边的图像显示了无线电波段的蟹状脉冲星 (Crab Pulsar) 图像，为了进行有意义的比较，还显示了使用前面讨论的一些波段拍摄的相同区域的图像。注意，每幅图像都给出了完全不同的脉冲星"视图"。

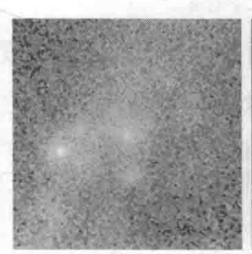
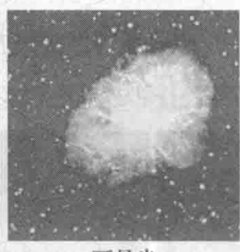
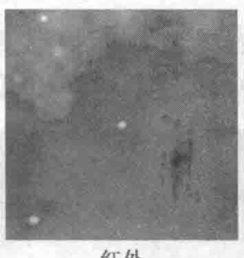
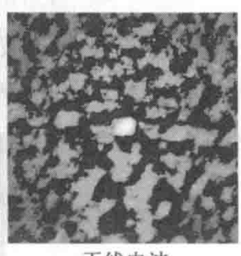





图 1.18　覆盖电磁波谱的 (位于图像中心的) 蟹状脉冲星图像 (图像由 NASA 提供)

1.3.7　使用其他成像方式的例子

虽然电磁波谱成像一直占主导地位，但大量的其他成像方式也很重要。特别地，本节讨论声波成像、电子显微方法和 (由计算机产生的) 合成成像。

利用"声音"成像在地质勘探、工业和医学中得到了应用。地质应用采用的是声谱中的低端声波 (几百赫兹)，其他应用领域的成像使用超声波 (百万赫兹)。图像处理在地质中的最重要商业应用是矿产和石油勘探。为了透过地表获取图像，主要方法之一是利用一辆大型卡车和一块大钢板，钢板由卡车压在地面上，同时卡车以 100 Hz 的频率振动。返回声波的强度和速度由地表下面的成分决定。这些声波经过计算机分析后，可由分析结果生成图像。

为了从海洋获取图像，能源通常由两个位于船后部的气枪组成。返回的声波由置于船后部电缆中的水听器检测，电缆要么放置在海底，要么用浮标悬吊 (垂直电缆)。两把气枪交替加压至约 2000 磅/平方英寸，然后快速释放压力。船的匀速运动提供横向运动，这种横向运动与返回的声波一起产生一幅海洋底部地下的合成三维图像。

图1.19 显示了一个著名三维模型的剖面图像，与此图像对照，测试地震成像算法的性能。箭头指向碳氢化合物 (油或气) 的油气阱。这一目标比周围地层明亮，因为目标区域内的密度变化较大。地震解释人员寻找这些亮点来发现油气。上面的地层也是亮的，但其亮度变化不像对面地层的亮度那样强。许多地震重建算法由于以上缺点使得对该目标成像比较困难。

42 ~ 43

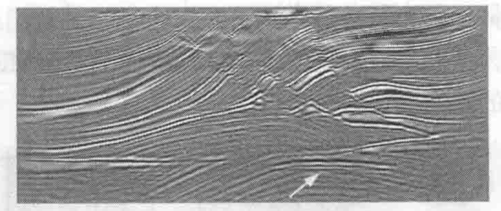


图 1.19　地震模型的剖面图像。箭头指向碳氢化合物（油或气）的油气阱(图像由美国 Sandia 国家实验室的 Curtis Ober 博士提供)

虽然超声波成像常用于制造业，但这一技术最为熟知的应用是在医学领域，特别是妇产科。在妇产科，医生对未出生的胎儿成像，以确定其发育的健康状况。这一检测的副产品是确定胎儿的性别。超声波图像是利用下面的基本步骤生成的：

1. 超声波系统(一台计算机、由超声波源和接收器组成的超声波探头及一台显示器)向人体发射高频(1~5 MHz)声波脉冲。
2. 声波传入体内并碰撞组织间的边界(举例来说，就是流体和软组织之间的边界，及软组织和骨骼之间的边界)。一部分声波反射回探头；一部分声波则继续传播，直到它们到达另一个边界并被反射。
3. 反射波被探头拾取并传给计算机。
4. 计算机根据声波在组织中的传播速度(1540 m/s)和每个回波的返回时间，计算从探头到组织或器官边界的距离。
5. 系统在屏幕上显示回波的距离和亮度，形成一幅二维图像。

在典型的超声波成像中，每秒有上百万个脉冲与回波发出和接收。探头可沿身体表面以不同角度运动，以便得到各种观察图像。图1.20显示了几个例子。

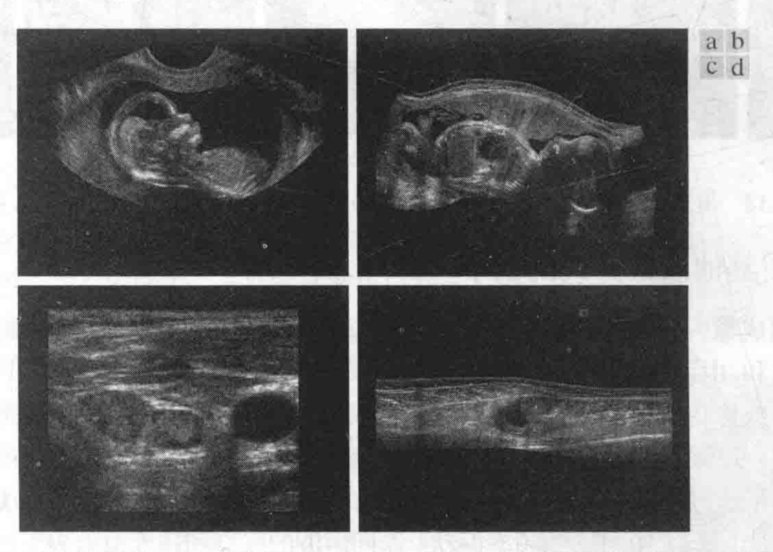


图 1.20　超声波成像示例：(a)胎儿图像；(b)胎儿的另一幅观察图像；(c)甲状腺图像；(d)有损伤的肌肉层图像(原图像由 Siemems Medical Systems 公司的 Ultrasound Group 提供)

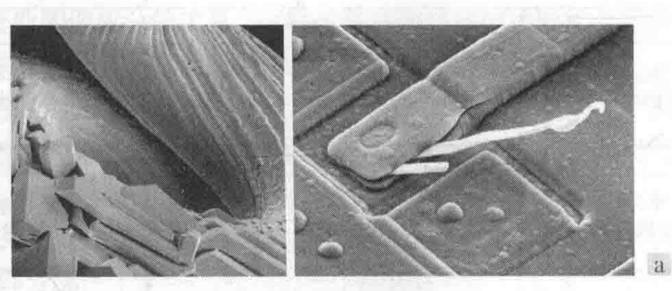
我们继续使用电子显微方法的某些例子来讨论成像方式。电子显微镜的功能与光学显微镜一样，只不过是用一个聚焦的电子波束代替光束形成图像标本。电子显微镜的操作包括下面几个基本步骤：由电子源产生一个电子流，用正电势朝着标本加速。该电子束用金属小孔和电磁透镜限制并聚焦为细的单色波束。然后使用一个磁透镜将单色电子束聚焦到样本上。在样本内部产生交互照射并影响电子束。这些交互作用和影响被检测并转换为一幅图像，这种方式与光被物体反射或吸收的方式一样。所有电子显微镜均执行这些基本步骤。

44
~
45

透射电子显微镜(TEM)的工作原理很像一台幻灯片投影仪。投影仪发射出一束透过幻灯片的光；

当光通过幻灯片时，它由幻灯片的内容调节。这一发射出的光束然后被投射到观察屏上，形成幻灯片的放大图像。除了发射的是通过样本(相当于幻灯片）的电子束外，TEM 以相同的方式工作。穿过样本的部分电子束被投射到荧光屏上，而电子与荧光物质的相互作用就产生了光，从而得到可观察的图像。另一方面，*扫描电子显微镜*(SEM) 实际上扫描电子束并记录每一位置上电子束与样本的相互作用。这会在荧光屏上产生一个点。一幅完整的图像由通过样本的电子束光栅扫描形成，这非常像电视摄像机。电子与荧光屏相互作用并产生光。SEM 适合于“大块”样品，而 TEM 要求非常薄的样品。

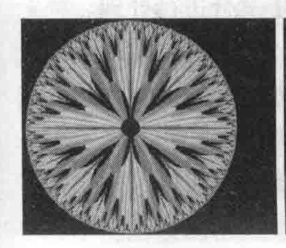
电子显微镜有非常高的放大能力。光学显微镜的放大倍数限制在 1000 左右，电子显微镜的放大倍数可达 10 000 或更大。图1.21 显示了两幅由于过热而造成损坏的样本图像。

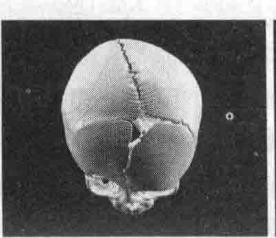
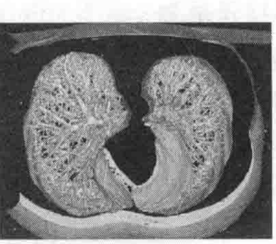


a b

图 1.21　(a) 因过热而损坏的钨丝放大 250 倍的 SEM 图像(注意左下方的碎片)；(b) 已损坏集成电路的 2500 倍 SEM 图像。白色纤维是由热破坏造成氧化的结果［图 (a) 由俄勒冈大学地质科学系的 Michael Shaffer 先生提供，图 (b) 由加拿大 McMaster 大学的 J. M. Hudak 博士提供］

下面通过简略浏览一些图像来结束对成像模式的讨论，这些图像不是来自物理对象，而是用计算机产生的。*分形*是由计算机生成图像的显著例子 (Lu[1997])。基本上，分形不过是一种根据某些数学规则进行迭代复制的基本模式。例如，“拼贴”是产生分形图像的最简方法之一。一个方块可细分为 4 个小方块，每个小方块又可进一步细分为 4 个更小的方块，等等。为了填充每个小方块，根据规则的复杂性，一些漂亮的瓷砖图像可用这一方法产生。当然，几何形状可以是任意的。例如，分形图像可以从一个中心点向外辐射生长。图1.22 (a) 显示了使用这种方法生成的分形图像。图1.22 (b) 显示了另一幅分形图像(“月球表面”)，它有趣地模拟了前述章节所用过的空间图像。



a b c d

图 1.22　(a)~(b) 分形图像；(c)~(d) 由所示物体的三维计算机模型生成的图像［图 (a) 和图 (b) 由 Swarthmore 学院的 Melissa D. Binde 先生提供，图 (c) 和图 (d) 由 NASA 提供］

分形图像由具有某种规则的子图像元素生长的数学公式表示，并朝艺术性方向发展。它们有时会用做随机纹理。借助于计算机生成更结构化的图像依赖于三维建模。三维建模是图像处理和计算机图形学的交叉领域，也是许多三维可视化系统(如光模拟器)的基础。图 1.22 (c) 和图 1.22 (d) 显示了计算机生成的图像例子。因为原始物体是以三维方式创建的，所以可以在任何角度由三维物体的平面投影来生成图像。这种类型的图像可用于医学培训及其他主要应用，如刑事法庭和需要产生特效的场合。

45 ~ 47

1.4 数字图像处理的基本步骤

把后续各章涉及的内容划分为 1.1 节定义的两个主要类别是有帮助的：一类是其输入和输出都是图像；另一类是其输入可能是图像但输出是从这些图像中提取的属性。这种组织结构总结于图1.23 中。该图并不意味着每种处理都用于图像。相反，目的是给出所有方法的概念，这些方法可以针对不同的目的、不同的目标运用于图像。这一节的讨论也可视为本书其余部分内容的综述。

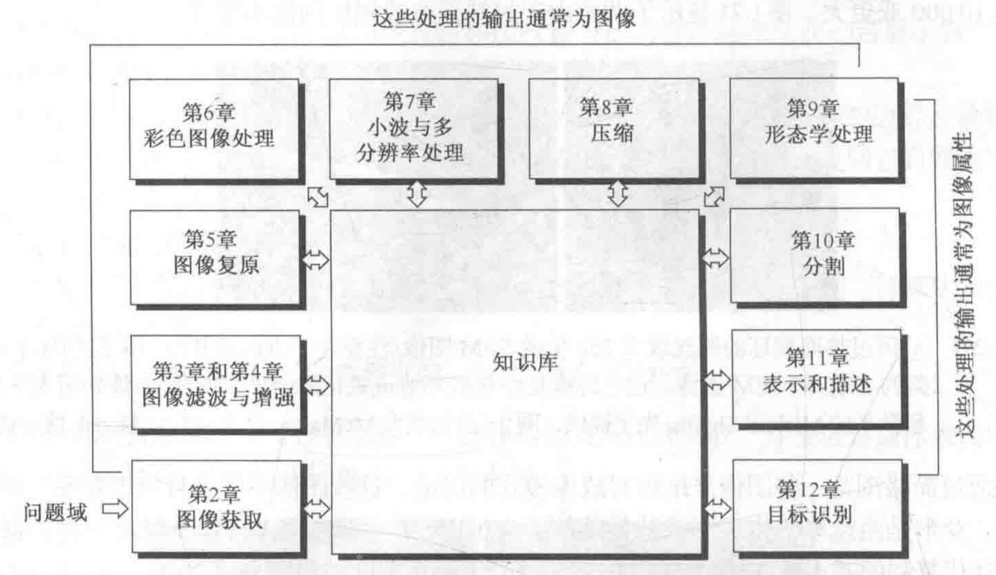


图 1.23 数字图像处理的基本步骤，方框中指出的章号是方框中讨论内容的位置

图像获取是图 1.23 中的第一步处理。1.3 节的讨论给出了关于数字图像起源的一些提示。这个主题在第 2 章中将更具体地进行研究，这一章还要介绍一些贯穿全书的基本数字图像的概念。注意，图像获取与给出一幅数字形式的图像一样简单。通常，图像获取阶段包括图像预处理，譬如图像缩放。

图像增强是对一幅图像进行某种操作，使其结果在特定应用中比原始图像更适合进行处理。特定一词在这里很重要，因为一开始增强技术就建立在面向问题的基础之上。例如，对增强 X 射线图像十分有用的方法，对增强电磁波谱中红外波段获取的卫星图像可能就不是最好的方法。

不存在增强图像的通用“理论”。当为视觉解释而处理一幅图像时，观察者就是特殊方法工作好坏的最终裁判者。增强技术是如此多种多样，以至于有众多不同的图像处理方法，因此在没有广泛背景介绍的情况下，在一章中组织出适合于增强技术的有意义主体是很困难的。基于这个原因，也基于图像处理领域的初学者通常看到图像增强应用在视觉上是很吸引人的、有趣且理解起来相对简单的原因，在第 2 章、第 3 章和第 4 章中介绍新概念时，我们都会以图像增强作为例子。后两章中的内容涉及许多用于增强图像的传统方法。因此，使用图像增强的例子来介绍前面章节阐述的新图像处理方法，不仅可节省本书处理图像增强的额外章节，而且更重要的是，可有效地向初学者介绍本书早期处理技术的细节。然而，如本书中其他内容所述的那样，这些章节阐述的内容可应用于比图像增强宽泛得多的问题。

图像复原也是改进图像外观的一个处理领域。然而，与图像增强不同，图像增强是主观的，而图像复原是客观的；在某种意义上说，复原技术倾向于以图像退化的数学或概率模型为基础。另一方面，增强以什么是好的增强效果这种主观偏爱为基础。

47 ~ 48

彩色图像处理已成为一个重要领域，因为互联网上数字图像的使用正在不断增长。第 6 章涵盖

了许多彩色模型和数字域彩色处理的基本概念。在后续各章中，彩色也是提取图像中感兴趣特征的基础。

小波是以不同分辨率来描述图像的基础。特别地，本书中为图像数据压缩和金字塔表示使用了小波，此时图像被成功地细分为较小的区域。

正如其名称指出的那样，压缩指的是减少图像存储量或降低传输图像带宽的处理。虽然存储技术在过去的 10 年里已有明显改进，但对于传输能力我们还不能这样说。互联网上的应用更是如此，因为互联网是以大量图片内容为特征的。多数计算机用户都很熟悉(也许并未注意)图像压缩所用的图像文件扩展名，如 JPG 文件扩展名用于 JPEG(联合图片专家组)图像压缩标准。

形态学处理涉及提取图像成分的工具，这些成分在表示和描述形状方面很有用。正如1.1 节所指出的那样，这一章的内容将从输出图像处理到输出图像属性处理的转换开始。

分割过程将一幅图像划分为其组成部分或目标。通常，自动分割是数字图像处理中最困难的任务之一。成功地把目标逐一识别出来是一个艰难的分割过程。另一方面，很弱且不稳定的分割算法几乎总是会导致最终失败。通常，分割越准确，识别越成功。

表示与描述几乎总是在分割阶段的输出之后，通常这一输出是未加工的像素数据，这些数据不是构成一个区域的边界(即分隔一个图像区域与另一个图像区域的像素集合)，就是构成该区域本身的所有点。无论哪种情况，把数据转换成适合计算机处理的形式都是必要的。首先，必须确定数据是应表示为一条边界还是应表示为整个区域。如果关注的是外部形状特征，譬如角点和拐点，则表示为边界是合适的。如果我们关注的是内部特性，例如纹理或骨架形状，则区域表示是合适的。在某些应用中，这些表示彼此是互补的。选择一种表示仅是把原始数据转换为适合计算机进行后续处理的形式的一部分。为了描述数据以使感兴趣的特征更明显，必须确定一种方法。描述又称为特征选择，它涉及提取特征，可得到某些感兴趣的定量信息，或是区分一组目标与其他目标的基础。

识别是基于目标的描述给该目标赋予标志(譬如“车辆”)的过程。如1.1 节详细介绍的那样，我们通过阐述识别个别目标的方法来结束数字图像处理的讨论。

到目前为止，我们还没有谈到关于先验知识及图1.23 中知识库与各个处理模块之间的关系。有关问题域的知识已经以知识库的形式编码并存入图像处理系统中。这一知识可能像一幅图像详细描述区域那样简单，在这里定位已知的感兴趣的信息，可将限制性的搜索引导到要寻找的信息处。知识库也可能相当复杂，如材料检测问题中所有主要缺陷的相关列表，或者包含变化检测应用中一个区域的高分辨率卫星图像的图像数据库。除了引导每个处理模块的操作外，知识库还要控制模块之间的交互。这一特性由图1.23 中处理模块和知识库之间的双头箭头表示，而单头箭头则用于连接处理模块。

虽然没有专门讨论图像显示，但要记住图像显示是很重要的，观察图像处理的结果可在图 1.23 中的任何阶段的输出处执行。还应注意，并非所有图像处理应用都需要图 1.23 给出的复杂交互。事实上，在某些情况下甚至并不需要所有这些模块。例如，用于人眼视觉解释的图像增强就很少要求使用图 1.23 中的任何其他步骤。然而，通常随着图像处理任务复杂性的增加，需要做更多的处理才能解决问题。

1.5 图像处理系统的组成

20 世纪 80 年代中叶，世界各地售出的各种型号的图像处理系统，基本上都是由许多主机及与这些主机配套的外设组成的。20 世纪 80 年代末 90 年代初，市场已转为将图像处理硬件设计为与工业标准总线兼容，并能配合工程工作站机箱和个人计算机的单板形式。除了降低成本外，这一市场转型还如催化剂一样催生了大量的新公司，这些公司的任务是开发用于图像处理的软件。

49 ~ 50

虽然针对如处理卫星图像的大规模图像应用的大型图像处理系统一直在出售，但趋势是朝着小型化和通用化的小型机并带有专用图像处理硬件的混合型系统方向发展。图1.24显示了用于数字图像处理的一个典型通用系统的组成。下面几段讨论每个组件的功能，从图像感知开始。

关于感知，需要两个部件来获取数字图像。第一个部件是物理设备，该设备对我们希望成像的目标辐射的能量很敏感。第二个部件称为数字化器，数字化器是一种把物理感知装置的输出转换为数字形式的设备。例如，在数字视频摄像机中，传感器产生一个与光强成正比的输出，数字化器把该输出转换为数字数据。这些主题包含在第2章中。

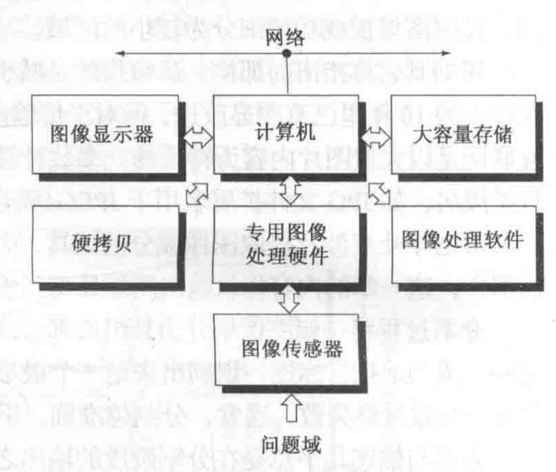


图1.24 通用图像处理系统的组成

专用图像处理硬件通常由刚刚谈到的数字化器与执行其他原始操作的硬件[如算术逻辑单元(ALU)]组成，算术逻辑单元对整个图像并行执行算术与逻辑运算。使用ALU的一个例子是与数字化一样快的图像平均操作，这一操作的目的是降低噪声。这种类型的硬件有时称为前端子系统，其显著特点是速度快。换句话说，该单元执行要求快速数据吞吐的功能(譬如以30帧/秒的速率来数字化和平均视频图像)，典型的主机不能胜任该工作。

图像处理系统中的计算机是通用计算机，其范围从PC到超级计算机。有时在专门应用中也采用特殊设计的计算机，以达到所要求的性能。但是，这里感兴趣的还是通用图像处理系统。在这些系统中，几乎任何配置较好的PC对于离线图像处理任务都是适合的。

图像处理软件由执行特定任务的专用模块组成。一个设计优良的软件包还包括为用户写代码的能力，例如，最小化就可以使用专用模块。更完善的软件包集成了这些模块，并至少能用一种计算机语言编写通用软件命令集。

大容量存储能力在图像处理应用中是必需的。尺寸为1024×1024像素的一幅图像，每像素的灰度是8比特，如果图像不压缩，则需要1 MB的存储空间。在处理几千幅甚至几百万幅图像时，在图像处理系统中提供足够的存储空间将是一个极大的挑战。图像处理应用的数字存储分为三个主要类别：(1)处理期间的短期存储；(2)关系到快速调用的在线存储；(3)档案存储，其特点是不频繁访问。存储是以字节(8比特)、千字节(1000字节)、兆字节(100万字节)、吉字节(10亿字节)或太字节(TB)来计量的。

提供短期存储的一种方法是使用计算机内存。另一种方法是采用专用的存储板，这种存储板称为帧缓存，它们可以存储一帧或多帧图像并可快速访问，访问速率通常为视频速率(30帧/秒)。后一种方法实质上允许瞬时缩放、滚动(垂直移动)和摇动(水平移动)图像。帧缓冲器通常放在专用图像处理硬件单元中，如图1.24所示。在线存储通常采用磁盘或光介质存储。在线存储的关键特性参数是对存储数据的访问频率。最后，档案存储是以大容量存储要求为特征的，但无须频繁访问。放在类似于投币电唱机机盒内的磁带和光盘是常用的档案存储介质。

今天使用的图像显示器主要是彩色电视监视器(更好一些的是平面屏幕)。监视器由图像和图形显示卡的输出驱动，它们是计算机系统的一个集成部分。对于图像显示应用不可能没有这样的要求，即作为计算机系统的一部分，其显示卡应满足商用性要求，在有些情况下还要求立体显示。立体显示是通过戴在用户头上的目镜上嵌入两个小显示屏的头盔来实现的。

用于记录图像的硬拷贝设备包括激光打印机、胶片相机、热敏装置、喷墨装置和数字单元，如CD-ROM等。胶片的分辨率最高，但纸作为书写材料是首选介质。为了表现图像，图像可显示在透明胶片上，或使用图像投影设备显示在数字介质中。后者作为图像表现的标准正在被人们逐步接受。

50 ~ 52

网络在今天所用的计算机系统中几乎都是默认的功能。因为在图像处理应用中会形成大量数据，图像传输中主要考虑的问题是带宽。在专用网络中，这不是什么大问题，但通过互联网进行远程通信并不总是有效的。所幸的是，随着光纤与其他宽带技术的发展，这一状况正在迅速得到改善。

小结

本章的主要目的是简要介绍数字图像处理的起源、重要意义、当前和未来的应用领域。由于篇幅所限，本章中涉及主题的覆盖面必然不太全面，但在数字图像处理的知识宽度和应用范围方面应能给读者留下深刻的印象。后续章节将介绍图像处理理论与应用，并给出大量实例。学完本书的最后一章后，读者应能了解数字图像处理领域的当前研究进展情况。

参考文献

后续各章末尾的参考资料都是针对所讨论的特定主题提供的，它们在本书末尾提供。但在本章中，我们先列出图像处理领域的主要杂志和参考书籍，以便读者能快速了解本领域的历史和当前的发展动态。本章引用的参考材料主要是图像处理领域的普通文献。

图像处理及相关领域的主要杂志包括：*IEEE Transactions on Image Processing*；*IEEE Transactions on Pattern Analysis and Machine Intelligence*；*Computer Vision, Graphics, and Image Processing*（1991年前）；*Computer Vision and Image Understanding*；*IEEE Transactions on Systems, Man and Cybernetics*；*Artificial Intelligence*；*Pattern Recognition*；*Pattern Recognition Letters*；*Journal of the Optical Society of America*（1984年前）；*Journal of the Optical Society of America-A: Optics, Image Science and Vision*；*Optical Engineering*；*Applied Optics Information Processing*；*IEEE Transaction on Medical Imaging*；*Journal of Electronic Imaging*；*IEEE Transactions on Information Theory*；*IEEE Transactions on Communications*；*IEEE Transactions on Acoustics, Speech and Signal Processing*；*Proceedings of the IEEE*；以及1980年前的*IEEE Transactions on Computer*专刊。国际光学工程协会(SPIE)的出版物也很重要。

下面按出版年份反序列出了包含有数字图像处理补充材料的书籍(偏重于近年来出版的书籍)。这些书籍给出了该领域过去30多年来的发展概述，同时选择了各种形式的论著。它们的覆盖范围从基础理论的教科书，到提供技术综述的手册，以及包含该领域当前研究水平的代表性材料的论文集。

Prince, J. L. and Links, J. M. [2006]. *Medical Imaging, Signals, and Systems*, Prentice Hall, Upper Saddle River, NJ.

Bezdek, J. C. et al. [2005]. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer, New York.

Davies, E. R. [2005]. *Machine Vision: Theory, Algorithms, Practicalities*, Morgan Kaufmann, San Francisco, CA.

Rangayyan, R. M. [2005]. *Biomedical Image Analysis*, CRC Press, Boca Raton, FL.

Umbaugh, S. E. [2005]. *Computer Imaging: Digital Image Analysis and Processing*, CRC Press, Boca Raton, FL.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. [2004]. *Digital Image Processing Using MATLAB*, Prentice Hall, Upper Saddle River, NJ.

Snyder, W. E. and Qi, Hairong [2004]. *Machine Vision*, Cambridge University Press, New York.

Klette, R. and Rosenfeld, A. [2004]. *Digital Geometry—Geometric Methods for Digital Picture Analysis*, Morgan Kaufmann, San Francisco, CA.

53~54

Won, C. S. and Gray, R. M. [2004]. *Stochastic Image Processing*, Kluwer Academic/Plenum Publishers, New York.

Soille, P. [2003]. *Morphological Image Analysis: Principles and Applications*, 2nd ed., Springer-Verlag, New York.

Dougherty, E. R. and Lotufo, R. A. [2003]. *Hands-on Morphological Image Processing*, SPIE—The International Society for Optical Engineering, Bellingham, WA.

Gonzalez, R. C. and Woods, R. E. [2002]. *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.

Forsyth, D. F. and Ponce, J. [2002]. *Computer Vision—A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.

Duda, R. O., Hart, P. E., and Stork, D. G. [2001]. *Pattern Classification*, 2nd ed., John Wiley & Sons, New York.

Pratt, W. K. [2001]. *Digital Image Processing*, 3rd ed., John Wiley & Sons, New York.

Ritter, G. X. and Wilson, J. N. [2001]. *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton, FL.

Shapiro, L. G. and Stockman, G. C. [2001]. *Computer Vision*, Prentice Hall, Upper Saddle River, NJ.

Dougherty, E. R. (ed.) [2000]. *Random Processes for Image and Signal Processing*, IEEE Press, New York.

Etienne, E. K. and Nachtegael, M. (eds.). [2000]. *Fuzzy Techniques in Image Processing*, Springer-Verlag, New York.

Goutsias, J., Vincent, L., and Bloomberg, D. S. (eds.). [2000]. *Mathematical Morphology and Its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Boston, MA.

Mallot, A. H. [2000]. *Computational Vision*, The MIT Press, Cambridge, MA.

Marchand-Maillet, S. and Sharaiha, Y. M. [2000]. *Binary Digital Image Processing: A Discrete Approach*, Academic Press, New York.

Mitra, S. K. and Sicuranza, G. L. (eds.) [2000]. *Nonlinear Image Processing*, Academic Press, New York.

Edelman, S. [1999]. *Representation and Recognition in Vision*, The MIT Press, Cambridge, MA.

Lillesand, T. M. and Kiefer, R. W. [1999]. *Remote Sensing and Image Interpretation*, John Wiley & Sons, New York.

Mather, P. M. [1999]. *Computer Processing of Remotely Sensed Images: An Introduction*, John Wiley & Sons, New York.

Petrou, M. and Bosdogianni, P. [1999]. *Image Processing: The Fundamentals*, John Wiley & Sons, UK.

Russ, J. C. [1999]. *The Image Processing Handbook*, 3rd ed., CRC Press, Boca Raton, FL.

Smirnov, A. [1999]. *Processing of Multidimensional Signals*, Springer-Verlag, New York.

Sonka, M., Hlavac, V., and Boyle, R. [1999]. *Image Processing, Analysis, and Computer Vision*, PWS Publishing, New York.

Haskell, B. G. and Netravali, A. N. [1997]. *Digital Pictures: Representation, Compression, and Standards*, Perseus Publishing, New York.

Jahne, B. [1997]. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, Springer-Verlag, New York.

Castleman, K. R. [1996]. *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.

Geladi, P. and Grahn, H. [1996]. *Multivariate Image Analysis*, John Wiley & Sons, New York.

Bracewell, R. N. [1995]. *Two-Dimensional Imaging*, Prentice Hall, Upper Saddle River, NJ.

54 ~ 55

Sid-Ahmed, M. A. [1995]. *Image Processing: Theory, Algorithms, and Architectures*, McGraw-Hill, New York.

Jain, R., Rangachar, K., and Schunk, B. [1995]. *Computer Vision*, McGraw-Hill, New York.

Mitiche, A. [1994]. *Computational Analysis of Visual Motion*, Perseus Publishing, New York.

Baxes, G. A. [1994]. *Digital Image Processing: Principles and Applications*, John Wiley & Sons, New York.

Gonzalez, R. C. and Woods, R. E. [1992]. *Digital Image Processing*, Addison-Wesley, Reading, MA.

Haralick, R. M. and Shapiro, L. G. [1992]. *Computer and Robot Vision*, vols. 1 & 2, Addison-Wesley, Reading, MA.

Pratt, W. K. [1991] *Digital Image Processing*, 2nd ed., Wiley-Interscience, New York.

Lim, J. S. [1990]. *Two-Dimensional Signal and Image Processing*, Prentice Hall, Upper Saddle River, NJ.

Jain, A. K. [1989]. *Fundamentals of Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ.

Schalkoff, R. J. [1989]. *Digital Image Processing and Computer Vision*, John Wiley & Sons, New York.

Giardina, C. R. and Dougherty, E. R. [1988]. *Morphological Methods in Image and Signal Processing*, Prentice Hall, Upper Saddle River, NJ.

Levine, M. D. [1985]. *Vision in Man and Machine*, McGraw-Hill, New York.

Serra, J. [1982]. *Image Analysis and Mathematical Morphology*, Academic Press, New York.

Ballard, D. H. and Brown, C. M. [1982]. *Computer Vision*, Prentice Hall, Upper Saddle River, NJ.

Fu, K. S. [1982]. *Syntactic Pattern Recognition and Applications*, Prentice Hall, Upper Saddle River, NJ.

Nevatia, R. [1982]. *Machine Perception*, Prentice Hall, Upper Saddle River, NJ.

Pavlidis, T. [1982]. *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, MD.

Rosenfeld, A. and Kak, A. C. [1982]. *Digital Picture Processing*, 2nd ed., vols. 1 & 2, Academic Press, New York.

Hall, E. L. [1979]. *Computer Image Processing and Recognition*, Academic Press, New York.

Gonzalez, R. C. and Thomason, M. G. [1978]. *Syntactic Pattern Recognition: An Introduction*, Addison-Wesley, Reading, MA.

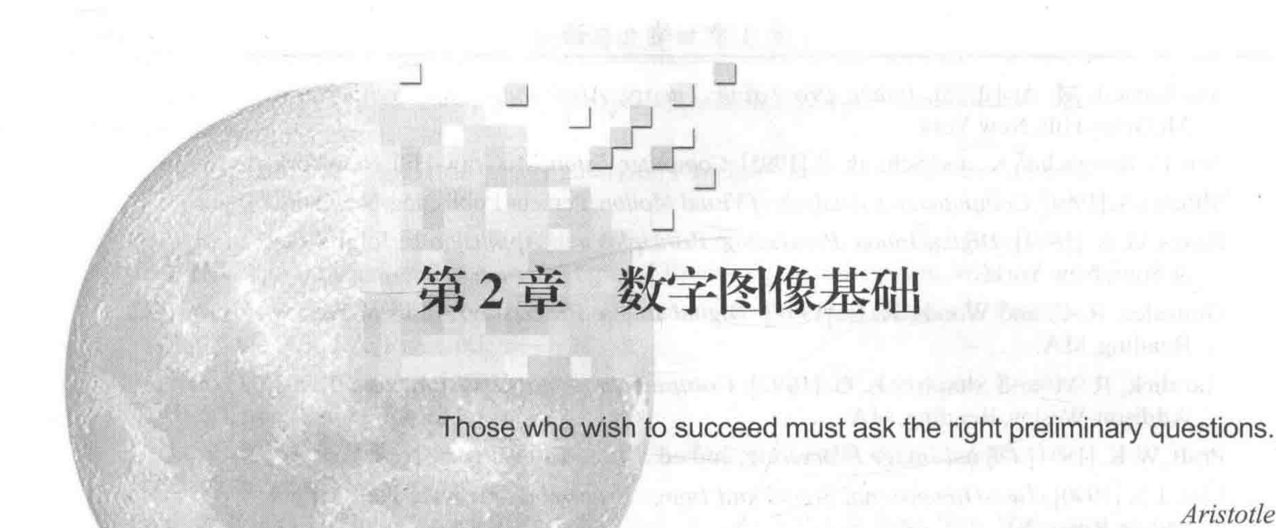
Andrews, H. C. and Hunt, B. R. [1977]. *Digital Image Restoration*, Prentice Hall, Upper Saddle River, NJ.

Pavlidis, T. [1977]. *Structural Pattern Recognition*, Springer-Verlag, New York.

Tou, J. T. and Gonzalez, R. C. [1974]. *Pattern Recognition Principles*, Addison-Wesley, Reading, MA.

Andrews, H. C. [1970]. *Computer Techniques in Image Processing*, Academic Press, New York.

55 ~ 56



第2章　数字图像基础

Those who wish to succeed must ask the right preliminary questions.

Aristotle

引言

本章的主要目的是介绍本书所用到的数字图像处理的一些基本概念。2.1 节简述人类视觉系统的机理，包括眼中图像的形成及对亮度的适应和鉴别能力。2.2 节讨论光、电磁波谱的其他成分及它们的成像特点。2.3 节讨论成像传感器及怎样使用它们产生数字图像。2.4 节介绍均匀图像取样及灰度量化的概念。该节中讨论的其他主题包括数字图像表示、图像中取样数和灰度级变化的影响、空间和灰度分辨率的概念，以及图像内插的原理。2.5 节介绍像素间的各种基本关系。最后，2.6 节介绍本书用到的主要数学工具。该节的第二个目的是帮助读者积累一些在各种基本图像处理任务中如何运用这些工具的经验。需要时，这些工具及其应用范围将在本书的剩余部分展开。

2.1　视觉感知要素

虽然数字图像处理这一领域建立在数学和概率公式表示的基础之上，但人的直觉和分析在选择一种技术而不选择另一种技术时会起核心作用，这种选择通常是基于主观的视觉判断做出的。因此，将大概了解人类的视觉感知作为学习本书的起点是合适的。鉴于这一主题的复杂性和宽泛性，这里仅涉及人类视觉的最基本方面。特别地，我们的兴趣在于人类形成并感知图像的机理和参数。我们的兴趣还在于通过数字图像处理的一些要素来了解人类视觉的物理限制。因此，人类与电子成像设备相比，在分辨率和对光照变化的适应能力方面不仅有趣，从实践的角度来看也很重要。

2.1.1　人眼的结构

图2.1 显示了人眼的一个简化剖面图。眼睛的形状近似为一个球体，其平均直径约为 20 mm。眼睛由三层膜包裹：角膜与巩膜外壳、脉络膜和视网膜。角膜是一种硬而透明的组织，覆盖着眼睛的前表面。与角膜相连的巩膜是一层包围眼球其余部分的不透明膜。

脉络膜位于巩膜的正下方。脉络膜包含有血管网，它是眼睛的重要滋养源。即使是对脉络膜表面并不严重的损害，也有可能严重地损害眼睛，引起限制血液流动的炎症。脉络膜外壳着色很重，因此有助于减少进入眼内的外来光和眼球内反向散射光的数量。脉络膜的最前面分为睫状体和虹膜。虹膜的收缩和扩张控制着进入眼睛的光量。虹膜中间开口(瞳孔)的直径可变，变化范围为 2~8 mm。

57
~
58

虹膜的前面包含有眼睛的可见色素，而后面则包含有黑色色素。

晶状体由同心纤维细胞层组成，并由附在睫状体上的纤维悬挂。晶状体包含60%~70%的水、6%的脂肪和比眼中任何其他组织都多的蛋白质。晶状体由稍黄的色素着色，其颜色随着人的年龄的增大而加深。在极端情况下，晶状体会过于混浊，这通常是由白内障等疾病引起的，可能导致低下的彩色辨别能力和视觉清晰度的损失。晶状体吸收大约8%的可见光谱，对短波长的光有较高的吸收率。在晶状体结构中，蛋白质吸收红外光和紫外光，吸收过量时会伤害眼睛。

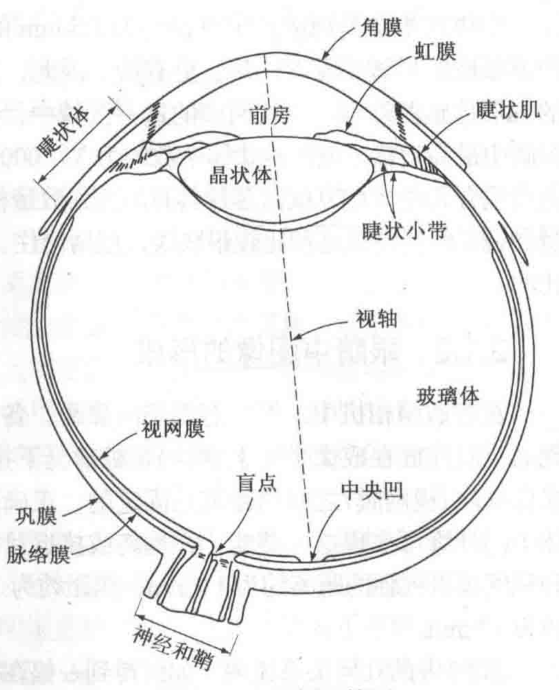


图2.1　人眼剖面简图

眼睛最里面的膜是视网膜，它布满了整个后部的内壁。眼睛适当地聚焦时，来自眼睛外部物体的光在视网膜上成像。感受器通过感受视网膜表面分布的不连续光来形成图案。光感受器分类两类：锥状体和杆状体。每只眼睛中的锥状体数量约为600~700万个。它们主要位于视网膜中称为中央凹的中间部分，对颜色高度敏感。使用这些锥状体，人们可以充分地分辨图像细节，因为每个锥状体都连接到了自身的神经末梢。肌肉控制眼球的转动，直到感兴趣物体图像落到中央凹上。锥状体视觉称为白昼视觉或亮视觉。

杆状体的数量更多：约有7500~15 000万个杆状体分布在视网膜表面。由于分布面积较大而且几个杆状体连接到一个神经末梢，因此减少了这些感受器感知细节的数量。杆状体用来给出视野内的总体图像。它们没有彩色感觉，对低照明度敏感。例如，白天色彩鲜艳的物体，在月光下却没有颜色，因为此时只有杆状体受到刺激。这种现象称为暗视觉或微光视觉。

图2.2显示了右眼中通过眼睛光神经出现区的剖面的杆状体和锥状体密度。在这一区域，由于没有感受器而导致了所谓的盲点(见图2.1)。除了这一区域，感受器的分布关于中央凹径向对称。感受器密度根据距中央凹的度数来度量(即离开视轴的度数，它由视轴和通过晶状体中心并相交于视网膜的一条直线的夹角度量)。注意，图2.2中的锥状体在视网膜的中心最密(在中央凹的中心区域)。还应注意到，从该中心向外到偏离视轴大约20°处，杆状体的密度逐渐增大，然后向外到视网膜的极限边缘处，密度逐渐下降。

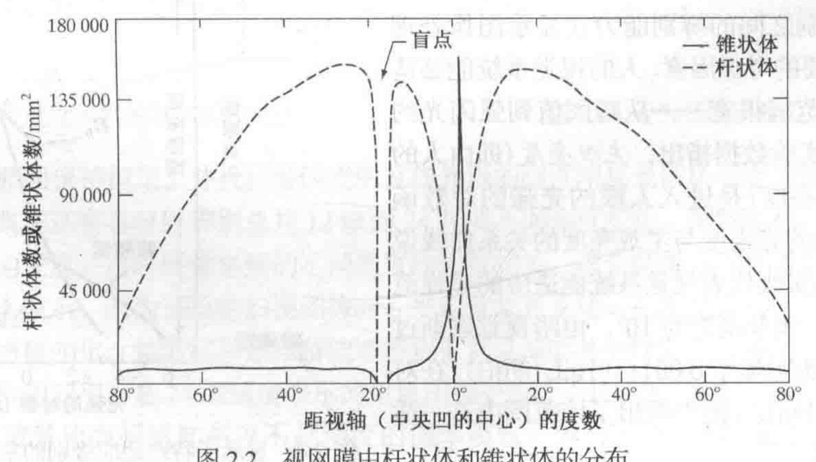


图2.2　视网膜中杆状体和锥状体的分布

58~60

中央凹本身是视网膜中直径约为 1.5 mm 的圆形凹坑。然而，为便于后续的讨论，我们把它近似为方形或矩形敏感元素阵列会更有用。因此，为解释方便，可把中央凹视为大小为 1.5 mm × 1.5 mm 的方形传感器阵列。在视网膜的这一区域中，锥状体的密度约为 150 000 个/ mm^2。基于这一近似，眼睛中最高敏感区域的锥状体个数约为 337 000。从自然分辨能力的角度看，恰好与一个中等分辨率的电荷耦合元件(CCD)成像芯片具有的元素数量相当，接收器阵列不大于 5 mm × 5 mm。尽管人类的智慧和视觉经验使得这种比较很肤浅，但请记住，人眼分辨细节的能力与当前电子成像传感器是可以类比的。

2.1.2 眼睛中图像的形成

在普通照相机中，镜头有固定的焦距，各种距离的聚焦是通过改变镜头和成像平面间的距离实现的，胶片放在成像平面上(数码相机情况下是成像芯片)。在人眼中，情形则与此相反；晶状体和成像区域(视网膜)之间的距离是固定的，正确聚焦的焦距是通过改变晶状体的形状来得到的。睫状体中的纤维可实现这一功能，在远离或接近目标物时，纤维会分别压扁或加厚晶状体。晶状体中心和视网膜沿视轴的距离约为 17 mm，焦距约为 14~17 mm。在眼睛放松且聚焦距离大于 3 m 时，焦距约为 17 mm。

图2.3 中的几何关系说明了如何得到一幅在视网膜上形成的图像的尺度。例如，假设一个人正在观看距其 100 m 处的高为 15 m 的一棵树。令 h 表示视网膜图像中该物体的高度，由图2.3 的几何形状可以看出 $15/100 = h/17$ 或 $h = 2.55$ mm。正如 2.1.1 节指出的那样，视网膜图像主要聚焦在中央凹区域。然后，光接收器的相对刺激作用产生感知，把辐射能转变为电脉冲，最后由大脑解码。

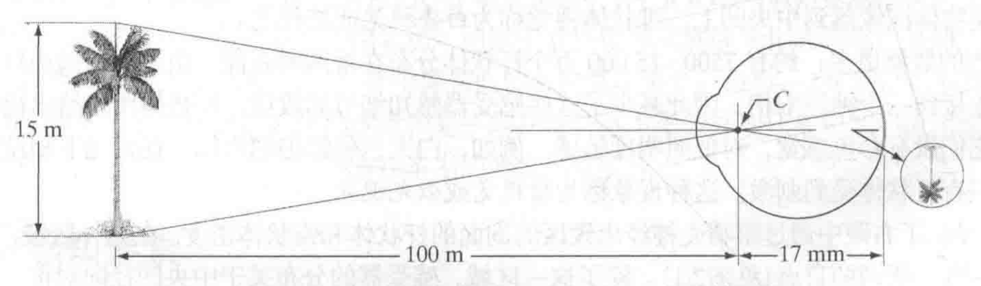


图 2.3 人眼观看一棵棕榈树的图解，点 C 是晶状体的光心

2.1.3 亮度适应和辨别

因为数字图像作为离散的灰度集来显示，所以眼睛对不同亮度级别之间的辨别能力在显示图像处理结果时是一个重要的考虑因素。人的视觉系统能够适应的光强度级别范围很宽——从暗阈值到强闪光约有 10^{10} 个量级。实验数据指出，主观亮度(即由人的视觉系统感知的亮度)是进入人眼的光强的对数函数。图 2.4 中画出的光强度与主观亮度的关系曲线说明了这一特性。长实线代表视觉系统能适应的光强范围。在亮视觉中，该范围约为 10^6。由暗视觉逐渐过渡到亮视觉的近似范围为 0.001~0.1mL(朗伯)(在对数坐标中为−3 ~ −1 mL)，图中画出了该范围内这一适应曲线的两个分支。

60 ~ 62

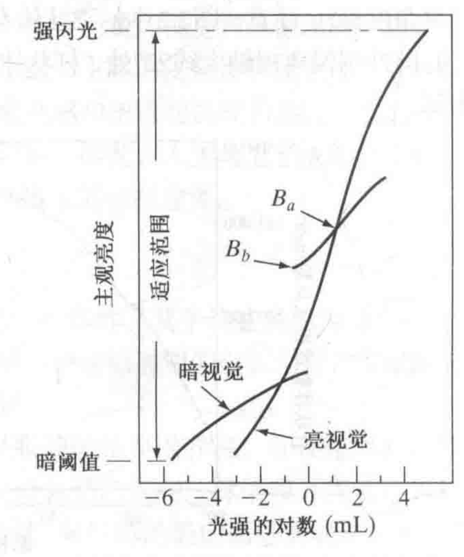


图 2.4 显示了特殊适应级别的主观亮度感知范围

解释图 2.4 中描述的令人印象深刻的动态范围的基本要点是，视觉系统不能同时在一个范围内工作。确切地说，它是通过改变其整个灵敏度来完成这一较大变动的，这就是周知的亮度适应现象。眼睛可以同时辨别的不同强度级别的总范围与整个适应范围相比很小。对于任何给定的条件集合，视觉系统的当前灵敏度级别称为*亮度适应级别*，例如，它可能对应于图 2.4 中的亮度 B_a。较短的交叉线表示当眼睛适应这一强度级别时，人眼所能感知的主观亮度范围。注意，这一范围是有一定限制的，级别 B_b 或低于级别 B_b 的刺激都感知为不可辨别的黑色。该曲线的上部实际上不受限制，但如果延伸太远也会失去意义，因为高得多的强度将会把适应能力提高到比 B_a 更高的水平。

在任何特定的适应级别上，人眼辨别光强变化的能力也是值得考虑的因素。用于确定人类视觉系统亮度辨别能力的一个著名实验，由一个注视对象和大到足以使其占有全部视野的均匀发光区组成。这一区域是一个典型的漫反射体，譬如不透明玻璃，使用强度 I 可变的一个光源从后面照向该漫反射体。在该视野中增加一个照射分量 ΔI，形成一个持续时间很短的闪烁，该闪烁以均匀光场中央的圆形方式出现，如图2.5所示。

如果 ΔI 不够亮，则目标不变，表明没有可觉察的变化。当 ΔI 逐渐加强时，目标会给出肯定的响应“是”，指出一个可觉察的变化。最后，当 ΔI 足够强时，目标将始终给出肯定的响应“是”。$\Delta I_c / I$ 称为*韦伯比*，其中 ΔI_c 是在背景照明为 I 时可辨别照明增量的 50%。$\Delta I_c / I$ 值较小意味着可辨别强度较小的百分比变化，这表示亮度辨别能力“较好”。相反，$\Delta I_c / I$ 较大意味着要求有较大强度的百分比变化，这表示亮度辨别能力“较差”。

作为 $\log I$ 的函数，$\log \Delta I_c / I$ 曲线具有图2.6所示的一般形状。这一曲线表明，在低照明级别，亮度辨别较差(韦伯比大)，且它会随着背景照明的增加而明显改善(韦伯比降低)。曲线中的两个分支反映了这样一个事实，即在低照明水平情况下，视觉由杆状体执行，在高照明水平情况下(表示较好的辨别能力)，视觉由锥状体执行。

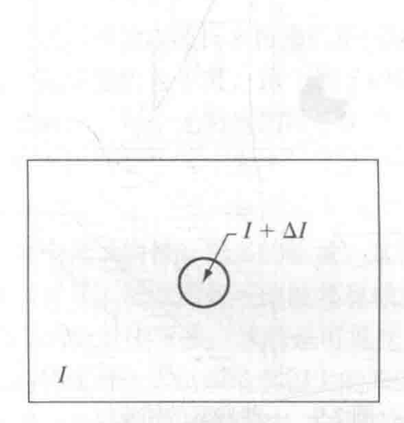


图 2.5　用于表征亮度辨别的基本实验设置

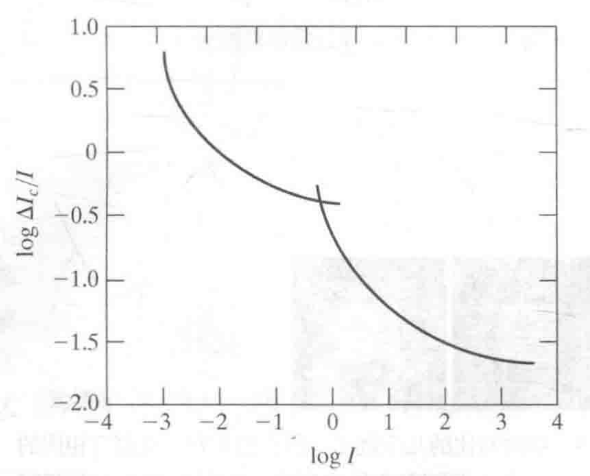


图 2.6　作为强度函数的典型韦伯比

如果背景照明保持恒定，并且代替闪光的其他光源的强度现在允许从不能觉察逐渐变化到总可以被觉察，那么典型观察者可以辨别总共 12 级到 24 级的不同强度变化。粗略地看，该结果与一个人观看单色图像中的任意一点时所觉察到的不同强度的数量相关。这个结果并不意味着一幅图像可以由这样少的强度值来表示，因为当眼睛扫视图像时，平均背景在变化，从而允许在每个新的适应级别上检测一组不同的增量变化。最终结果是眼睛能够辨别更宽范围的整个强度。事实上，我们将在2.4.3 节说明，眼睛能够检测到用不足 24 级强度表示的单色图像中的令人讨厌的轮廓效应。

62
~
63

两种现象清楚地表明感知亮度不是强度的简单函数。第一种现象基于这样一个事实，即视觉

系统往往会在不同强度区域的边界处出现“下冲”或“上冲”现象。图2.7(a)显示了这种现象的一个典型例子。虽然条带的强度恒定，但在靠近边界处我们实际上感知到了带有毛边的亮度模式［见图2.7(c)］。这些看起来带有毛边的带称为马赫带，厄恩斯特·马赫于1865年首次描述了这一现象。

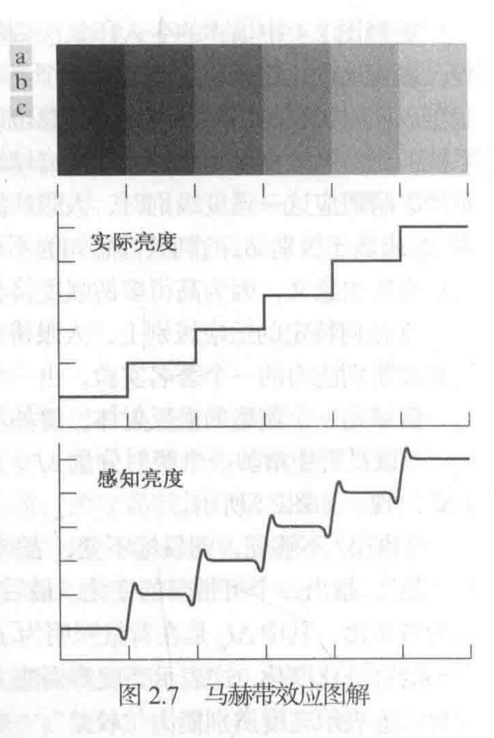


图2.7　马赫带效应图解

第二种现象称为同时对比，它与这样一个事实有关，即感知区域的亮度并不简单地取决于其强度，如图2.8所示。所有的中心方块都有完全相同的强度。然而，随着背景变得更亮，它们在眼睛里会变得更暗。一个更熟悉的例子是一张纸，一张纸放在桌子上时，看上去似乎比较白，而当我们使用它遮住眼睛来直视明亮的天空时，它看起来会总是呈黑色。

人类感知现象的另一些例子就是错觉，在错觉中，眼睛填充了不存在的信息或错误地感知了物体的几何特点。图2.9给出了一些例子。在图2.9(a)中，正方形的轮廓看起来很清楚，尽管图像中并没有定义这样一个图形的直线。在图2.9(b)中可以看到相同的效果，只是这次是一个圆；注意，仅有几条直线就足以导致一个完整圆的错觉。图2.9(c)中的两条水平线段的长度相同，但看起来一条显得比另一条短。最后，图2.9(d)中45°方向的所有直线都是等间距的平行线，然而画有交叉影线时就产生了错觉，觉得这些线不再平行。错觉是人类视觉系统的一种特性，但这一特性尚未被人类完全了解。

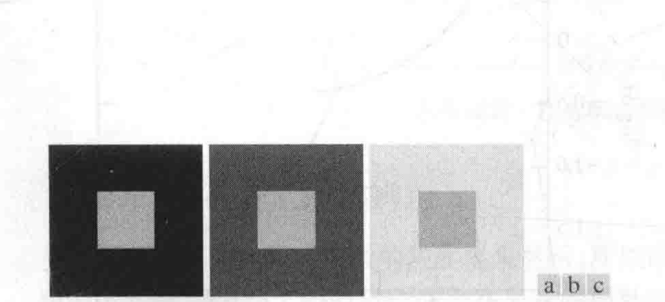


图2.8　同时对比的几个例子。所有的中心方块都有相同的强度，但随着背景的变亮，它们看起来逐渐变暗

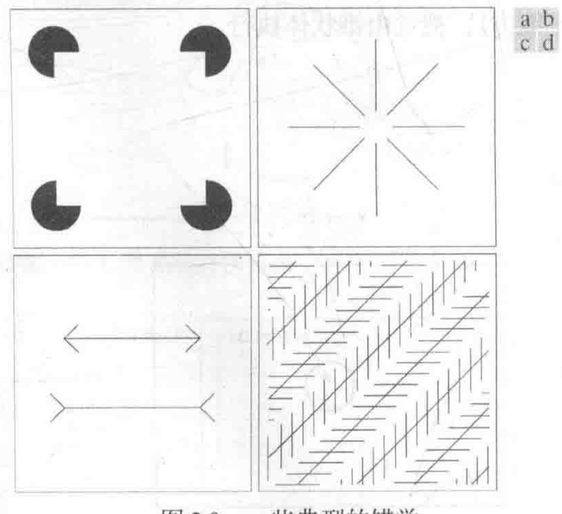


图2.9　一些典型的错觉

2.2　光和电磁波谱

电磁波谱在1.3节中已做过介绍，现在更详细地研究这一主题。1666年，艾萨克·牛顿发现，当一束太阳光通过一个玻璃棱镜后，显示的光束不再是白光，而是由一端为紫色而另一端为红色的连续色谱组成。如图2.10所示，我们感受到的可见光的彩色范围只占电磁波的一小部分。在波谱的一端是无线电波，其波长是可见光波长的几十亿倍。波谱的另一端是伽马射线，其波长比可见光小几百万倍。电磁波谱可用波长、频率或能量来描述。波长(λ)和频率(ν)的关系可用下式描述：

63
~
65

$$\lambda = c / v \tag{2.2-1}$$

式中 c 是光速(2.998×10^8 m/s)。电磁波谱的各个分量的能量由下式给出：

$$E = hv \tag{2.2-2}$$

式中 h 是普朗克常数。波长的单位是米(m)，最常用的单位是微米(表示为 μm，$1\ \mu m = 10^{-6}$ m)和纳米(表示为 nm，$1\ nm = 10^{-9}$ m)。频率用赫兹(Hz)来度量，1 Hz 表示正弦波每秒 1 个周期。常用的能量单位是电子伏特。

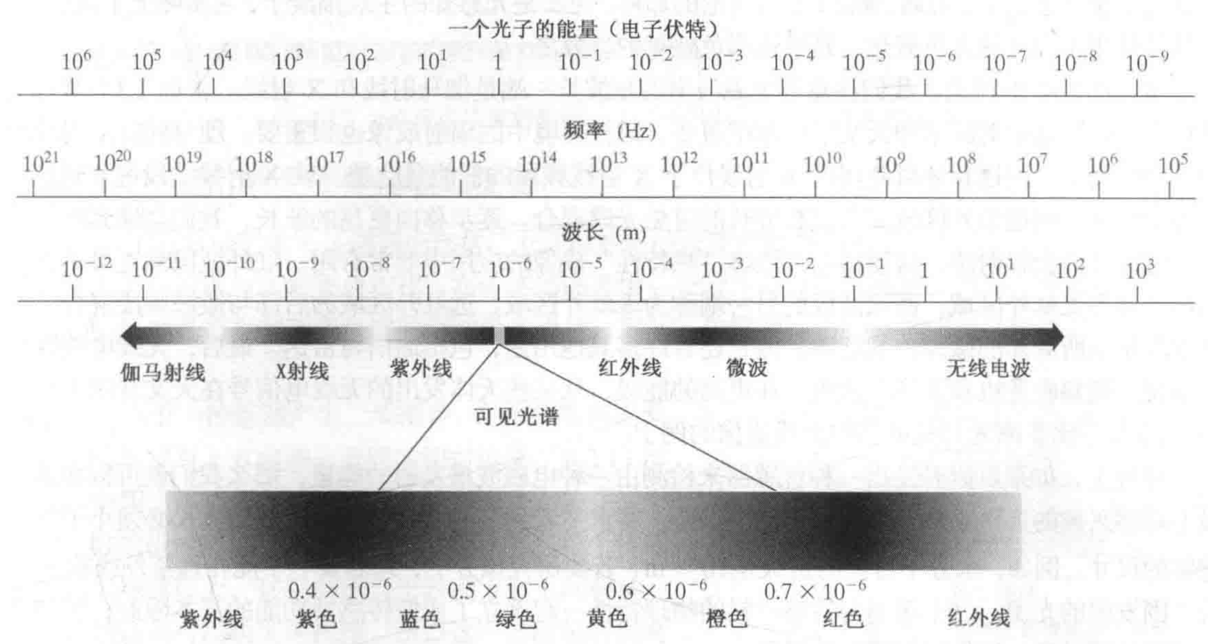


图 2.10　电磁波谱。为便于解释，可见光谱已被放大，但请注意，可见光谱是电磁波谱中相当窄的一部分

电磁波可视为以波长 λ 传播的正弦波(见图 2.11)，或视为没有质量的粒子流，每个粒子以波的模式以光速传播和移动。每个无质量的粒子包含一定的(一束)能量，每束能量称为一个光子。从式(2.2-2)可以看出能量与频率成正比，因此更高频率(更短波长)的电磁现象的每个光子携带有更多的能量。这样，无线电波有低能量光子，微波比无线电波具有较多的能量，红外波具有的能量还要多，然后是可见光、紫外线、X 射线和伽马射线，其中伽马射线的能量最高，这就是伽马射线对活体组织危害很大的原因。

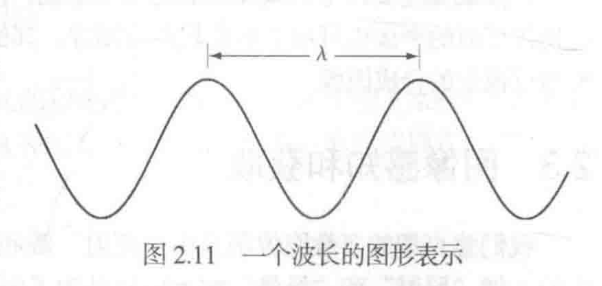


图 2.11　一个波长的图形表示

光是一种特殊的电磁辐射，它可以被人眼感知。为便于讨论，扩展后的可见光谱(彩色)段如图 2.10 所示(第 6 章将更详细地讨论彩色)。电磁波谱可见光波段的跨越范围为 0.43 μm(紫色)~0.79 μm(红色)。为方便起见，彩色谱分为 6 个主要区域：紫色、蓝色、绿色、黄色、橘黄色和红色。每种颜色(或电磁波谱的其他分量)不是突然终止的，而是混合平滑地过渡到另一种颜色的，如图 2.10 所示。

人感受到的物体颜色由物体反射的光的性质决定。以所有可见波长相对平衡地反射光的物体，对观察者而言是白色的。然而，一个物体在可见光谱的有限范围内反射时，会呈现各种颜色色调。例如，绿色物体反射波长范围为 500~570 nm 的光，而吸收其他波长的大部分能量。

没有颜色的光称为单色光或无色光。单色光的唯一属性是其强度或大小。因为感知单色光的强度从黑色到灰色变化，最后到白色，灰度级一词通常用来表示单色光的强度。在后续讨论中，我们

65~67

将交替地使用强度和灰度级这两个术语。从黑到白的单色光的度量值范围通常称为灰度级，而单色图像常称为灰度图像。

正如前面指出的那样，彩色光的电磁能谱的跨越范围为0.43 μm（紫色）~0.79 μm（红色）。除了频率之外，我们还用三个基本量来描述彩色光源的质量：发光强度、光通量和亮度。发光强度是从光源流出的能量总量，通常用瓦特（W）来度量。用流明数（lm）度量的光通量给出了观察者从光源感受到的能量。例如，从远红外光谱范围的光源发射出的光具有实际意义的能量，但观察者却很难感知到它。它的光通量几乎是零。最后，如2.1节所讨论的那样，亮度是光感知的主观描绘子，它实际上不能度量。它具体体现了强度的无色概念，是描述彩色感觉的参数之一。

继续图2.10的讨论，我们注意到电磁波谱的短波长一端是伽马射线和X射线。正如1.3.1节讨论的那样，伽马辐射对医学和天文学成像很重要，对核环境中的辐射成像也很重要。硬（高能）X射线用于工业应用中。胸透和牙科使用的X射线位于X射线频段的低能（软）端。软X射线波段过渡到远紫外线的区域，而远紫外区域又与较长波长的可见光谱混合。逐步移向更高的波长，我们会遇到红外波段，红外波段会辐射热，这使得它在依靠"热特性"成像的应用中非常有用。红外波段靠近可见光谱的部分称为近红外区域，而该波段的另一端称为远红外区域。远红外区域的后部与微波频段混合。这个波段是众所周知的微波炉的能源，但它还有许多其他用途，包括通信与雷达。最后，无线电波段包括电视、调幅收音机和调频收音机。在更高的能级，从某些天体发出的无线电信号在天文观测中很有用。1.3节给出了刚才讨论的多数波段成像的例子。

原理上，如果可以开发出一种传感器来检测由一种电磁波谱发射的能量，那么我们就可以在该波段上对感兴趣的事件成像。但要注意的一点是，要求"看到"一个物体的电磁波的波长必须小于等于物体的尺寸。例如，水分子直径的量级是10^{-10} m，若要研究该分子，则需要一个能在远紫外或软X射线范围发射的光源。这个限制与传感材料的物理特性一起确立了成像传感器功能的基本限制，譬如今天所用的可见光、红外线和其他传感器。

虽然成像主要以电磁波发射的能量为基础，但这并不是生成图像的唯一方法。例如，1.3.7节讨论的物体反射的声波也可用于形成超声波图像。其他的主要数字图像源是电子显微镜的电子束和用于图形与可视化的合成图像。

2.3 图像感知和获取

我们感兴趣的多数图像都是由"照射"源和形成图像的"场景"元素对光能的反射或吸收而产生的。把"照射"和"场景"加上引号是为了强调这样一个事实，即比我们所熟悉的一个可见光源每天照射普通的三维场景情况更一般。例如，照射可能由电磁能源引起，如雷达、红外线或X射线系统。但是，正如前面指出的那样，照射也可以由非传统光源（如超声波）甚至由计算机产生的照射模式产生。类似地，场景元素可能是熟悉的物体，但它们也可能是分子、沉积岩或人类的大脑。依赖光源的特性，照射被物体反射或透射。第一类例子是从平坦表面反射。第二类例子是为了产生一幅X射线照片，让X射线透过病人的身体。在某些应用中，反射能或透射能可聚焦到一个光转换器上（如荧光屏），光转换器再把能量转换为可见光。电子显微镜和某些伽马成像应用就使用这种方法。

图2.12显示了用于将照射能量变换为数字图像的三种主要传感器配置。原理很简单：通过组合输入电能和对特殊类型检测能源敏感的传感器材料，把输入能源转变为电压。输出电压波形是传感器的响应，通过把传感器响应数字化，从每个传感器得到一个数字量。在这一节中，我们将关注图像感知和生成的主要方式。图像数字化将在2.4节讨论。

67 ~ 69

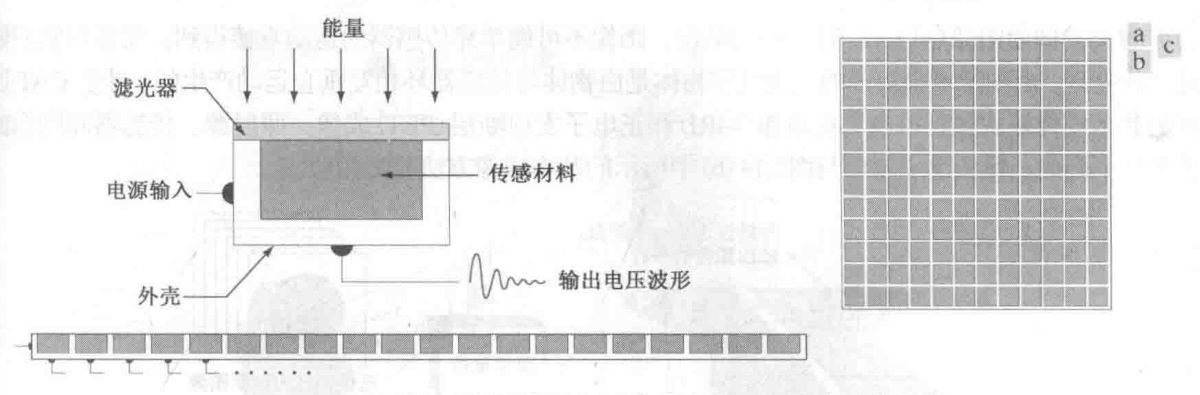


图 2.12　(a)单个成像传感器；(b)条带传感器；(c)阵列传感器

2.3.1　使用单个传感器获取图像

图2.12(a)显示了单个传感器的部件。也许最熟悉的这类传感器是光二极管，它由硅材料构成，并且其输出电压波形与入射光成正比。在传感器前面用一个滤光器改善选择性。例如，光传感器前面的绿色(通过)滤光器有利于彩色谱的绿波段光通过。因此，传感器输出的绿光比可见光谱中的其他成分要强。

为使用单一传感器产生二维图像，在传感器和成像区域之间必须有 x 方向和 y 方向的相对位移。图 2.13 示出了一个用于高精度扫描的配置，其中底片安装在一个鼓上，鼓的机械转动提供了一个维度的位移。单个传感器安装引导螺杆上，它提供垂直于转动方向上的移动。因为机械运动可高精度地控制，所以这一方法是得到高分辨率图像的一种廉价方法(但速度较慢)。另一种类似的机械配置使用一个平面床，传感器则在两个方向线性移动。这些类型的机械数字化仪有时称为微密度计。

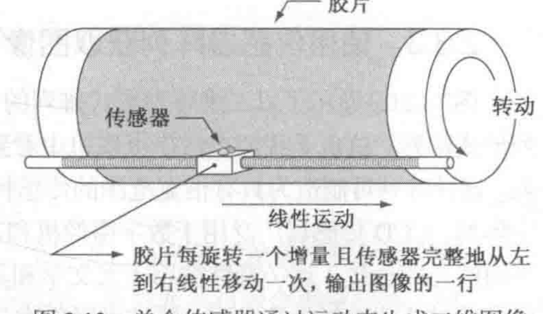


图 2.13　单个传感器通过运动来生成二维图像

使用单个传感器成像的另一个例子是将一个激光源和传感器放在一起。使用镜子来控制扫描模式的发射光束，并将反射的激光信号引导到传感器。这种配置也可使用条形或矩形传感器获取图像，这种配置将在后续两节中加以讨论。

2.3.2　使用条带传感器获取图像

与单个传感器相比，更常用的几何结构是由内嵌式传感器组成的传感器带，如图2.12(b)所示。这种传感器带在一个方向上提供成像单元。垂直于传感器带的运动在另一方向上成像，如图2.14(a)所示。这是大多数平板扫描仪中使用的排列方式。感知设备可能内嵌有 4000 个或更多的传感器。内嵌传感器常用于航空成像应用中，在这种应用中，成像系统安装在一架飞行器上，飞行器以恒定高度和速度飞过被成像的地区。能响应各种电磁波谱波段的一维传感器带垂直于飞行方向安装。成像传感器带一次给出一幅图像的一行，传感器带的运动则完成另一维度的二维图像。透镜和其他聚焦方法用于把扫描区域投影到传感器上。

以圆环形方式安装的传感器带用于医学和工业成像，以得到三维物体的剖面(“切片”)图像，如图2.14(b)所示。旋转的 X 射线源提供照射，射线源对面的传感器则收集穿过物体的 X 射线能量(很明显，这些传感器必须对 X 射线敏感)。这就是 1.2 节和 1.3.2 节中所述医学和工业计算机轴向断层(CAT)成像的基础。注意，传感器的输出必须由重建算法处理，重建算法的目的是把感知数据转换

69～71

为有意义的剖面图像(见5.11节)。换句话说，图像不可能单靠传感器的运动直接得到；需要对它们做进一步处理。由图像堆叠组成的三维数字物体是由物体与传感器环相互垂直运动产生的。基于CAT原理的其他成像模式包括核磁共振成像(MRI)和正电子发射断层(PET)成像。照射源、传感器和图像的类型是不同的，但概念上它们与图2.14(b)中所示的基本成像方法非常相似。

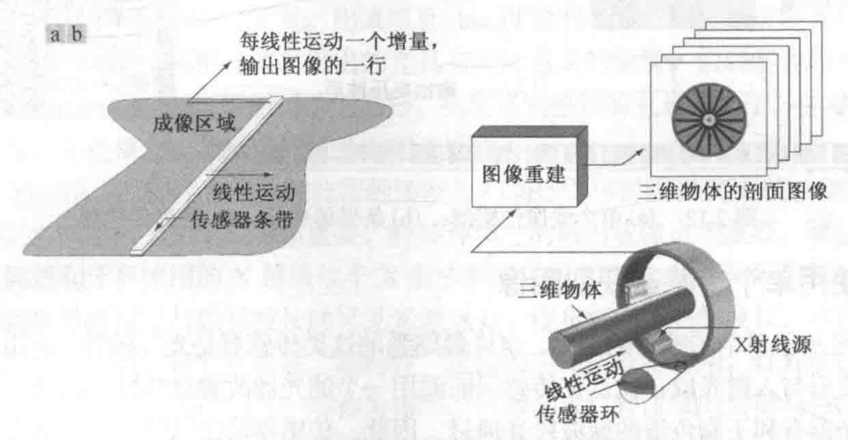


图2.14 (a)使用一个线性传感器带获取图像；(b)使用一个环形传感器带获取图像

2.3.3 使用传感器阵列获取图像

图2.12(c)显示了以二维阵列形式排列的一个传感器。大量的电磁波和一些超声波传感装置常以阵列形式排列。这也是我们在数字摄像机中看到的主要排列方式。这些摄像机的典型传感器是CCD阵列，这种阵列可制造为具有很宽范围的传感特性，并能封装为具有4000×4000个单元或更多单元的稳定阵列。CCD传感器广泛用于数字摄像机和其他光敏设备中。每个传感器的响应正比于投射到传感器表面的光能总量，这一特性被用于天文学和其他要求低噪声图像的应用中。通过让传感器累积输入光信号超过几分钟甚至几小时，可达到降低噪声的目的。因为图2.12(c)所示的传感器阵列是二维的，故其主要优点是通过将能量聚焦到阵列表面，来得到一幅完整的图像。很明显，前面讨论的传感器排列的运动是不需要的。

这样的传感器阵列所用的主要方法示于图2.15。该图显示来自照射源的能量是场景元素的反射(如本节开始提到的那样，该能量也可以通过场景元素透射)。图2.15(c)所示成像系统执行的第一个功能是收集入射能量，并将它聚焦到一个图像平面上。如果照射的是光，则成像系统的前端是一个光学透镜，该透镜把观察的场景投影到透镜的聚焦平面上，如图2.15(d)所示。与焦平面重合的传感器阵列产生与每个传感器接收到的光总量成正比的输出。数字或模拟电路扫描这些输出，并把它们转换成模拟信号，然后由成像系统的其他部分数字化，输出是一幅数字图像，如图2.15(e)所示。把一幅图像转换为数字形式是2.4节的主题。

> 在某些情况下，我们直接对源成像，就像得到太阳的图像一样。

2.3.4 简单的图像形成模型

正像在1.1节中介绍的那样，我们用形如$f(x, y)$的二维函数来表示图像。在空间坐标(x, y)处，f的值或幅度是一个正的标量，其物理意义由图像源决定。当一幅图像由物理过程产生时，其亮度值正比于物理源(如电磁波)所辐射的能量。因此，$f(x, y)$一定是非零的和有限的，即

$$0 < f(x, y) < \infty \tag{2.3-1}$$

> 在处理过程中或是作为解译的结果，图像灰度可以为负值。例如，在雷达成像中，移向雷达系统的物体通常解释为速度为负，而远离雷达系统的物体总是解释为速度为正。这样，一幅速度图像就可编码为既有正值又有负值。在存储和显示图像时，我们通常会标定灰度，以便最小的负值变为0(灰度标定将在2.6.3节介绍)。

71 ~ 73

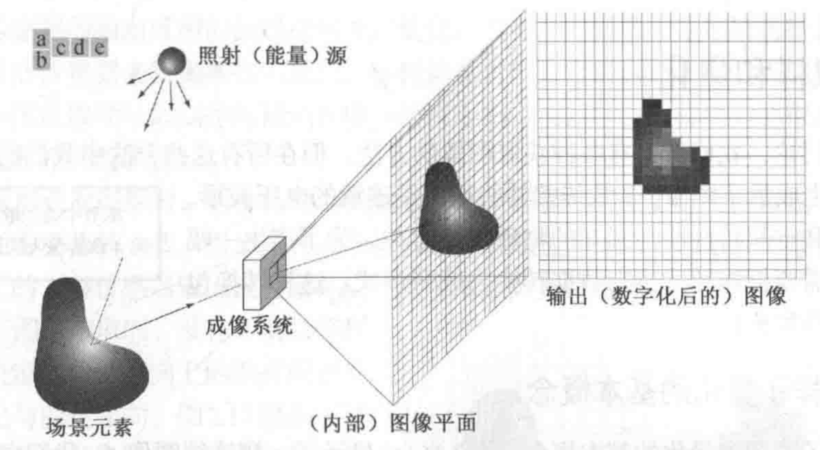


图 2.15　数字图像获取过程的一个例子：(a)能量（“照射”）源；(b)场景元素；(c)成像系统；(d)场景到图像平面的投影；(e)数字化后的图像

函数 $f(x,y)$ 可由两个分量来表征：(1)入射到被观察场景的光源照射总量；(2)场景中物体所反射的光照总量。这两个分量分别称为*入射分量*和*反射分量*，且分别表示为 $i(x,y)$ 和 $r(x,y)$ 。两个函数的乘积就是 $f(x,y)$，即

$$f(x,y)=i(x,y)r(x,y) \tag{2.3-2}$$

式中，

$$0<i(x,y)<\infty \tag{2.3-3}$$

和

$$0<r(x,y)<1 \tag{2.3-4}$$

式(2.3-4)指出反射分量限制在 0(全吸收)和 1(全反射)之间。$i(x,y)$ 的性质取决于照射源，而 $r(x,y)$ 的性质则取决于成像物体的特性。注意，这种表示方式还可用于照射光通过媒介形成图像的情况，如胸透 X 射线照片。在这种情况下，我们应该用*透射系数*代替*反射函数*，但其限制应该与式(2.3-4)相同，且形成的图像函数会建模为式(2.3-2)中的乘积。

例 2.1　照射和反射的一些典型值。

式(2.3-3)和式(2.3-4)中给出的值是理论界限。对于可见光，下面的平均数值用数字说明了 $i(x,y)$ 的一些典型范围。在晴朗的白天，太阳在地面上可能会产生超过 90 000 lm/m^2 的照度。在有云的白天，这一数值下降到 10 000 lm/m^2。在晴朗的夜晚，满月情况下的照度约为 0.1 lm/m^2。商用办公室的典型照度约为 1000 lm/m^2。类似地，下面是 $r(x,y)$ 的一些典型值：黑天鹅绒为 0.01，不锈钢为 0.65，白色墙为 0.80，镀银金属为 0.90，雪为 0.93。

令单色图像在任何坐标 (x_0, y_0) 处的强度(灰度)表示为

$$\ell=f(x_0,y_0) \tag{2.3-5}$$

由式(2.3-2)到式(2.3-4)可知 ℓ 的取值范围为

$$L_{\min}\leqslant\ell\leqslant L_{\max} \tag{2.3-6}$$

理论上，仅要求 $L_{\min}$ 为正，而要求 $L_{\max}$ 为有限值。实际上，$L_{\min}=i_{\min}r_{\min}$ 和 $L_{\max}=i_{\max}r_{\max}$ 。参照前面提到的办公室的平均照度和反射值范围，我们希望 $L_{\min}\approx10$ 和 $L_{\max}\approx1000$ 作为没有附加照明的室内值的典型限制。

区间 $[L_{\min}, L_{\max}]$ 称为*灰度级*(或*强度级*)。实际情况下常令该区间为 $[0, L-1]$，其中 $\ell=0$ 为黑色，$\ell=L-1$ 在灰度级中为白色。所有中间值是从黑色到白色之间变化的灰度色调。

73
~
74

2.4　图像取样和量化

由前一节的讨论，我们看到有多种获取图像的方法，但在所有这些方法中我们的目的是相同的，即从感知的数据生成数字图像。多数传感器的输出是连续的电压波形，这些波形的幅度和空间特性都与感知的物理现象有关。为了产生一幅数字图像，我们需要把连续的感知数据转换为数字形式。这种转换包括两种处理：取样和量化。

> 本节中关于取样的讨论较为直观。第4章将深入探讨该主题。

2.4.1　取样和量化的基本概念

图2.16说明了取样和量化的基本概念。图2.16(a)显示了一幅连续图像 f，我们想把它转换为数字形式。一幅图像的 x 和 y 坐标及幅度可能都是连续的。为将它转换为数字形式，必须对坐标和幅度都进行取样操作。对坐标值进行数字化称为取样，对幅值数字化称为量化。

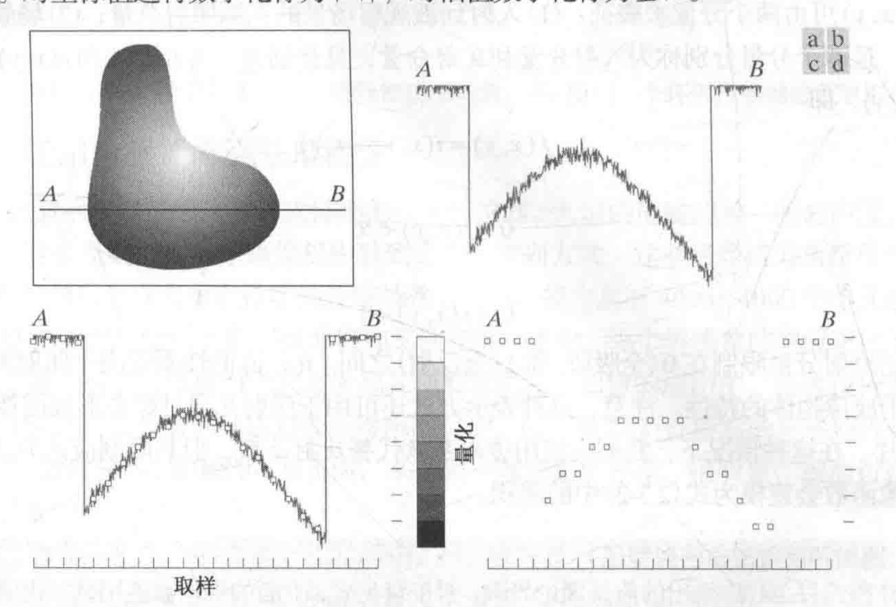


图2.16　生成一幅数字图像：(a)连续图像；(b)连续图像中从 A 到 B 的一条扫描线，用于说明取样和量化的概念；(c)取样和量化；(d)数字扫描线

图 2.16(b)中的一维函数是图 2.16(a)中沿线段 AB 的连续图像幅度值(灰度级)的曲线。随机变化是由图像噪声引起的。为了对该函数取样，我们沿线段 AB 等间隔地对该函数取样，如图 2.16(c)所示。每个样本的空间位置由图形底部的垂直刻度指出。样本用函数曲线上的白色小方块表示。这样的一组离散位置就给出了取样函数。然而，样本值仍(垂直)跨越了灰度值的连续范围。为了形成数字函数，灰度值也必须转换(量化)为离散量。图 2.16(c)的右侧显示了已分为 8 个离散区间的灰度标尺，范围从黑到白。垂直刻度标记指出了赋予 8 个灰度的每个特定值。通过对每个样本赋予 8 个离散灰度级中的一个来量化连续灰度级。赋值取决于该样本与一个垂直刻度标记的垂直接近程度。取样和量化操作生成的数字样本如图 2.16(d)所示。从该图像的顶部开始逐行执行这一过程，会产生一幅二维数字图像。图 2.16 意味着除了所用的离散级数外，量化所达到的精度强烈地依赖于取样信号的噪声。

74~75

按刚才讨论的方式取样，假设我们有一幅在两个坐标方向上和幅度上都连续的图像。实践中，取样方法由用于生成该图像的传感器配置决定。如图2.13所示，当一幅图像由单个传感单元与机械运动

相结合生成时，传感器的输出可用前面讨论的方式量化。然而，空间取样是通过选择各个机械增量的数量并在取样处激活传感器来收集数据完成的。机械运动可以非常精确地实现，因此从原理上讲，使用这种方法对一幅图像取样可以到达的精细程度并没有限制。实践中，取样精度的限制由其他因素决定，譬如系统的光学元件的质量。

使用带状传感器获取图像时，带中传感器的数量决定了图像中一个方向上的限制。另一个方向上的机械运动可控制得更精确一些，但试图在一个方向上实现超过另一个方向上被传感器数量限制的取样密度是没有意义的。对传感器输出进行量化后，就完成了产生数字图像的过程。

传感阵列用于图像获取时，没有运动且阵列中传感器的数量决定了两个方向上的取样限制。传感器输出的量化与前述相同。图2.17说明了这个概念。图2.17(a)显示了投影到一个阵列传感器平面上的连续图像。图2.17(b)显示了取样和量化后的图像。很明显，数字图像的质量在很大程度上取决于取样和量化中所用的样本数和灰度级。然而，如 2.4.3 节所述，在选择这些参数时，图像内容是一个重要的考虑因素。

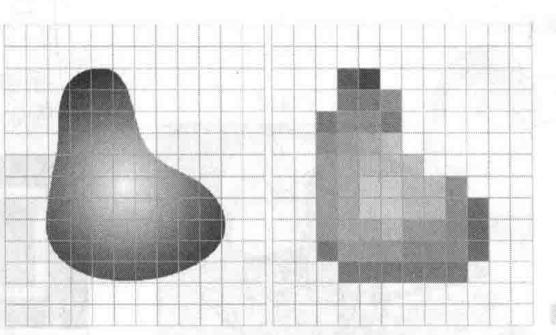


图 2.17　(a)已投影到一个传感器阵列上的连续图像；(b)图像取样和量化后的结果

2.4.2　数字图像表示

令 $f(s,t)$ 表示一幅具有两个连续变量 s 和 t 的连续图像函数。如前一节解释的那样，通过取样和量化，我们可把该函数转换为数字图像。假设我们把这幅连续图像取样为一个二维阵列 $f(x,y)$，该阵列包含有 M 行和 N 列，其中 (x,y) 是离散坐标。为表达清楚和方便起见，我们对这些离散坐标使用整数值：$x=0,1,2,\cdots,M-1$ 和 $y=0,1,2,\cdots,N-1$。这样，数字图像在原点的值就是 $f(0,0)$，第一行中下一个坐标处的值是 $f(0,1)$。这里，符号 $(0,1)$ 表示第一行的第二个样本，它并不是对图像取样时的物理坐标值。通常，图像在任何坐标 (x,y) 处的值记为 $f(x,y)$，其中 x 和 y 都是整数。由一幅图像的坐标张成的实平面部分称为空间域，x 和 y 称为空间变量或空间坐标。

如图2.18所示，有三种基本方法表示 $f(x, y)$。图2.18(a)是一幅函数图，它用两个坐标轴决定空间位置，第三个坐标是以两个空间变量 x 和 y 为函数的 f(灰度)值。虽然我们可以在这个例子中用该图来推断图像的结构，但通常复杂的图像细节太多，以至于很难由这样的图像去解译。在处理的元素是以 (x, y, z) 三坐标的形式表达的灰度集时，这种表示是很有用的，其中 x 和 y 是空间坐标，z 是 f 在坐标 (x, y) 处的值。在 2.6.4 节中，我们将探讨这种表示。

图 2.18(b)是更一般的表示。它显示了 $f(x,y)$ 出现在监视器或照片上的情况。这里，每个点的灰度与该点处的 f 值成正比。图中仅有三个等间隔的灰度值。如果灰度被归一化到区间[0, 1]内，那么图像中每个点的灰度都有 0, 0.5 或 1 这样的值。监视器或打印机简单地把这三个值分别变换为黑色、灰色或白色，如图 2.18(b)所示。第三种表示是将 $f(x,y)$ 的数值简单地显示为一个阵列(矩阵)。在这个例子中，f 的大小为 600×600 个元素，或 360 000 个数字。很清楚，打印整个矩阵是很麻烦的，且传达的信息也不多。然而，在开发算法时，当图像的一部分被打印并作为数值进行分析时，这种表示相当有用。图2.18(c)以图形方式传达了这一概念。

由前面几节的讨论，我们可得出一个结论，即图 2.18(b)和图 2.18(c)中的表示是最有用的。图像显示允许我们快速地观察结果。数值阵列用于处理和算法开发。以公式形式，我们可将一个 $M\times N$ 的数值阵列表示为

76 ~ 78

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix} \tag{2.4-1}$$

该式的两边以等效的方式定量地表达了一幅数字图像。右边是一个实数矩阵，矩阵中的每个元素称为图像单元、图像元素或像素。图像和像素这两个术语将在全书中表示数字图像及其元素。

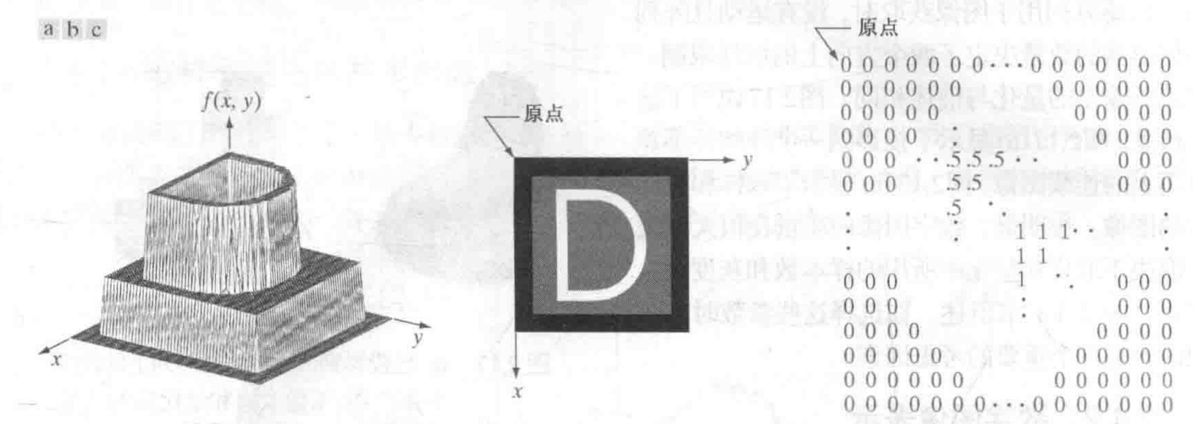


图 2.18 (a)画为表面图形的图像；(b)显示为可视灰度阵列的图像；(c)显示为一个二维数值阵列的图像(0, 0.5 和 1 分别表示黑色、灰色和白色)

在某些讨论中，使用传统的矩阵表示法来表示数字图像及其像素更为方便：

$$\boldsymbol{A}=\begin{bmatrix} a_{0,0} & a_{0,1} & \cdots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \cdots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \cdots & a_{M-1,N-1} \end{bmatrix} \tag{2.4-2}$$

显然，$a_{ij}=f(x=i,y=j)=f(i,j)$，因此式(2.4-1)和式(2.4-2)是相同的矩阵。我们甚至可将一幅图像表示一个向量 $\boldsymbol{v}$。例如，尺寸为 $MN\times1$ 的列向量由 $\boldsymbol{v}$ 的前 M 个元素作为 $\boldsymbol{A}$ 的第一列来构成，其后的 M 个元素作为第二列，等等。我们也可使用 $\boldsymbol{A}$ 的行代替列来形成这样的一个向量。只要一致，哪种表示都是有效的。

简要地回顾图2.18，注意到，数字图像的原点位于左上角，其中正 x 轴向下延伸，正 y 轴向右延伸。这种方便的表示基于如下事实：许多图像显示(譬如电视显示器)扫描都是从左上角开始的，然后一次向下移动一行。更重要的事实是，矩阵的第一个元素按惯例应在阵列的左上角，因此将 $f(x,y)$ 的原点选择在左上角数学上行得通。记住，这种表示是我们所熟悉的标准右手笛卡儿坐标系统①。我们仅说明了指向下方和指向右方的坐标轴来代替向右和向上的坐标轴。

有时，以更正式的数学术语表达取样和量化可能会很有用。令 Z 和 R 分别表示整数集和实数集。取样处理可视为把 xy 平面分为网格的过程，网格中每个单元的中心坐标是笛卡儿积 Z^2 中的一对元素，Z^2 是所有有序元素对 (z_i,z_j) 的集合，z_i 和 z_j 是 Z 中的整数。因此，如果 (x,y) 是 Z^2 中的整数，且 f 是把灰度值(即实数集 R 中的一个实数)赋给每个特定坐标对 (x,y) 的一个函数，则 $f(x,y)$ 就是

78 ~ 79

① 回忆可知，右手坐标系统是这样一种系统：当正 x 轴方向向右时，在大拇指竖直向上时，正 y 轴的方向为垂直于 x 轴的中指方向。如图 2.18(a)所示，事实上这是我们的图像坐标系统情形。

一幅数字图像。显然，这种赋值过程就是前面描述的量化处理。如果灰度级也是整数(此时及后续章节中通常为这种情况)，则用 Z 代替 R，然后，数字图像变成一个二维函数，且其坐标和幅值都是整数。

> 通常，将 L 个灰度值缩放至区间[0, 1]对于计算或算法开发目的是有用的，这时，它们不再是整数。但多数情形下，这些值都会被缩放到用于图像存储和显示的整数区间[0, L−1]。

数字化过程要求针对 M 值、N 值和离散灰度级数 L 做出判定。对于 M 和 N，除了必须取正整数外没有其他限制。然而，出于存储和量化硬件的考虑，灰度级数通常取为 2 的整数次幂，即

$$L = 2^k \tag{2.4-3}$$

我们假设离散灰度级是等间隔的，且它们是区间[0, L−1]内的整数。有时，灰度跨越的值域非正式地称为动态范围。这一术语在不同的场合有不同的用法。这里，我们将图像系统的动态范围定义为系统中最大可度量灰度与最小可检测灰度之比。原则上，上限取决于饱和度，下限取决于噪声(见图 2.19)。基本上，动态范围由系统能表示的最低和最高灰度级来确定，因此，这也是图像具有的动态范围。与这一概念紧密联系的是图像对比度，我们将一幅图像中最高和最低灰度级间的灰度差定义为对比度。当一幅图像中像素可感知的数值具有高动态范围时，那么我们认为该图像具有高对比度。相反，具有低动态范围的图像很沉闷，看上去灰度被冲淡。第 3 章将更加详细地讨论这些概念。

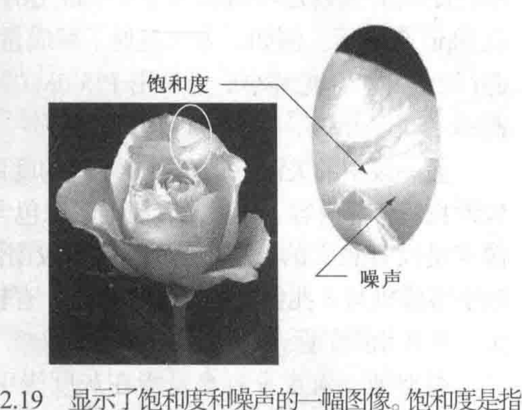


图 2.19　显示了饱和度和噪声的一幅图像。饱和度是指超过这个值的灰度级将被剪切掉的一个最高值(注意整个饱和区域具有恒定的高灰度级)。这种情况下的噪声表现为粒状纹理模式。噪声，特别是较暗图像区域(譬如玫瑰的茎)中的噪声，掩盖了可检测的最低真实灰度级

存储数字图像所需的比特数 b 为

$$b = M \times N \times k \tag{2.4-4}$$

当 $M = N$ 时，该式变为

$$b = N^2 k \tag{2.4-5}$$

表 2.1 中显示了 N 和 k 取不同值时需要用来存储方形图像的比特数。括号中的数字表示对应于每个 k 值的灰度级数。当一幅图像有 2^k 个灰度级时，实际上通常称该图像为一幅"k 比特图像"。例如，有 256 个可能的离散灰度值的图像称为 8 比特图像。注意，大小为1024×1024 或更高的 8 比特图像的存储需求很大。

表 2.1　N 和 k 取不同值时存储所需的比特数。L 是灰度级数

N/k	1 (L = 2)	2 (L = 4)	3 (L = 8)	4 (L = 16)	5 (L = 32)	6 (L = 64)	7 (L = 128)	8 (L = 256)
32	1 024	2 048	3 072	4 096	5 120	6 144	7 168	8 192
64	4 096	8 192	12 288	16 384	20 480	24 576	28 672	32 768
128	16 384	32 768	49 152	65 536	81 920	98 304	114 688	131 072
256	65 536	131 072	196 608	262 144	327 680	393 216	458 752	524 288
512	262 144	524 288	786 432	1 048 576	1 310 720	1 572 864	1 835 008	2 097 152
1024	1 048 576	2 097 152	3 145 728	4 194 304	5 242 880	6 291 456	7 340 032	8 388 608
2048	4 194 304	8 388 608	12 582 912	16 777 216	20 971 520	25 165 824	29 369 128	33 554 432
4096	16 777 216	33 554 432	50,331 648	67,108,864	83 886 080	100 663 296	117 440 512	134 217 728
8192	67 108 864	134 217 728	201 326 592	268,435,456	335 544 320	402 653 184	469 762 048	536 870 912

79~81

2.4.3 空间和灰度分辨率

直观上看，空间分辨率是图像中可辨别的最小细节的度量。在数量上，空间分辨率可以有很多方法来说明，其中*单位距离的线对数*和*单位距离的点数*(像素数)是最通用的度量。假设我们用交替出现的黑色和白色垂线来构造一幅图形，其中线宽为 W 个单位(W 可以小于 1)。线对的宽度是 $2W$，单位距离内有 $1/2W$ 个线对。例如，如果一条线的宽度是 0.1 mm，那么单位距离(mm)内就有 5 个线对。广泛使用的图像分辨率的定义是单位距离内可分辨的最大线对数量(譬如每毫米 100 个线对)。单位距离的点数是印刷和出版业中常用的图像分辨率的度量。在美国，这一度量通常使用*每英寸点数*(dpi)来表示。例如，为大到处了解质量的概念，报纸用 75 dpi 的分辨率来印刷，杂志则采用 133 dpi 的分辨率，光鲜小册子采用 175 dpi 的分辨率，我们正在阅读的书页则是以 2044 dpi 的分辨率印刷的。

前一段中的关键点是空间分辨率的度量必须针对空间单位来规定才有意义。图像大小本身并不会告诉我们全部内容。如果没有规定图像包含的空间维数，那么我们说一幅图像的分辨率为1024×1024像素是没有意义的。尺寸本身只在比较图像容量时才有帮助。例如，带有 20 兆像素成像芯片的 CCD 数字摄像机与 8 兆像素的摄像机相比，有较高分辨细节的能力，假定两部摄像机都配备了可比较的镜头，并在相同的距离拍摄可比较的图像。

类似地，*灰度分辨率*是指在灰度级中可分辨的最小变化。我们有关于产生数字图像的样本数的判断，但对于灰度级来说这并不适用。基于硬件考虑，如前节所述，灰度级数通常是 2 的整数次幂。最通用的数是 8 比特，在某些特殊的图像增强应用中，用 16 比特也是必要的。灰度量化用 32 比特则很罕见。有时，我们会发现使用 10 比特或 12 比特来数字化图像灰度级的系统，但这些系统都是特例而非常规系统。不像空间分辨率必须以单位距离为基础才有意义，灰度分辨率指的则是用于量化灰度的比特数。例如，通常说一幅被量化为 256 级的图像有 8 比特的灰度分辨率。因为灰度中可分辨的真实变化不仅受噪声和饱和度值的影响，也受人类感知能力的影响(见 2.1 节)，因此 8 比特灰度分辨率的一幅图像所含的内容，并不比用固定灰度值增量 1/256 来量化为 8 比特的系统更多。

下面的两个例子分别说明了图像尺寸和灰度分辨率对分辨细节的影响。在本节稍后，我们将讨论决定感知图像质量中的两个参数是如何相互影响的。

例 2.2 降低图像空间分辨率的效果说明。

图 2.20 显示了降低一幅图像的空间分辨率的效果。图 2.20(a)到(d)显示了分辨率分别为 1250 dpi, 300 dpi, 150 dpi 和 72 dpi 的图像。显然，低分辨率的图像与原图像相比要小。例如，原图像的大小为3692×2812像素，但 72 dpi 图像是一个大小为213×162 的阵列。为便于细节的比较，所有小图像都放大到了原图像大小(采用 2.4.4 节中讨论的放大方法)。

图 2.20(a)和图 2.20(b)之间有一些较小的视觉差别，最明显的差别是大黑针稍微有点儿失真。然而，图2.20(b)的大部分还是可以接受的。事实上，300 dpi 是书籍印刷所用的最小空间分辨率。因此，这里我们不能期望看到很大的不同。图2.20(c)显示了可见的退化(例如，计时器的圆形边缘和右侧指向 60 处的小针)。图2.20(d)显示了图像中多数可见特征的退化。如 4.5.4 节中讨论的那样，当以这样低的分辨率印刷时，印刷和出版业会使用一些“技巧”(例如，像素尺寸的局部变化)来产生比图2.20(d)更好的结果。此外，如 2.4.4 节所示，选择所用的内插方法也可以改进图 2.20 的结果。

81
~
82

例 2.3 改变数字图像中灰度级数的典型效果。

在这个例子中，我们保持样本数恒定，而将灰度级数以 2 的整数次幂从 256 减少到 2。图2.21(a)是一

幅大小为 452×374 的 CT 投影图像，它以 $k = 8$(256 个灰度级)显示。这样的图像是通过将 X 射线源固定在一个位置，然后在任何期望的方向生成二维图像来得到的。投影图像用于指导设置 CT 扫描仪的参数，包括倾角、切片数和范围。

图2.21(b)到图2.21(h)是在保持图像大小为 452×374 像素不变的情况下，比特数从 $k = 7$ 减小到 $k = 1$ 时得到的图像。256 级、128 级和 64 级灰度图像对于所有实用目的在视觉上的效果是相同的。然而，图 2.21(d)所示的 32 灰度级图像中，在恒定或接近恒度灰度区域(尤其在头盖骨处)内有一组不易察觉的细小脊状结构。这种效果是由数字图像的平滑区域中的灰度级数不足引起的，通常称为伪轮廓，之所以这样称呼，是因为这些脊状结构类似于地图中的地形轮廓。伪轮廓通常在以 16 或更少级数均匀设置的灰度级显示的图像中十分明显，如图2.21(e)到图2.21(h)所示的图像。

为方便起见，通常假设数值为 2 的整数次幂，一般来说，尺寸为 256×256 像素、64 灰度级、印刷在 5 cm × 5 cm 版式上的图像，是具有最低空间和灰度分辨率的图像，没有令人讨厌的取样棋盘格和伪轮廓。

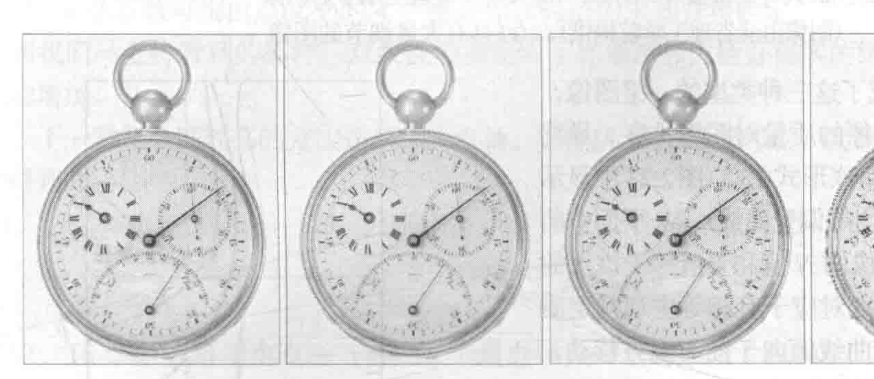


图 2.20　降低空间分辨率的典型效果。图像分别以(a) 1250 dpi，(b) 300 dpi，(c) 150 dpi，(d) 72 dpi 来显示。为清楚起见，图像添加了黑色的细边框，它们不是数据的一部分

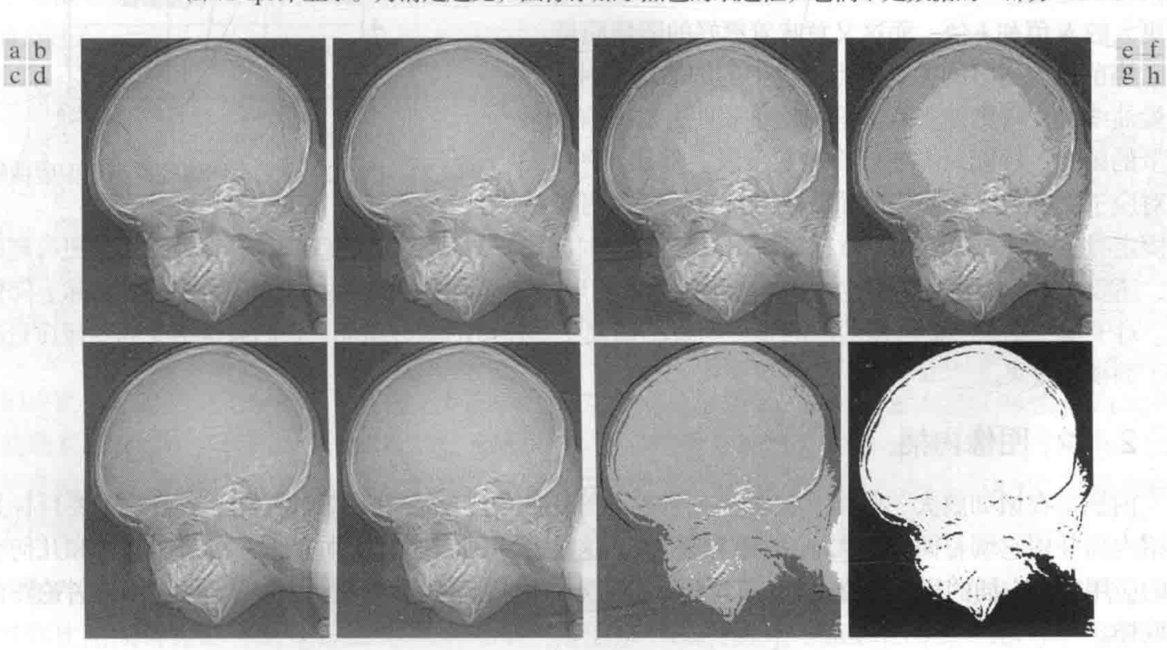


图 2.21　(a)大小为 452×374 的 256 灰度级图像；(b)~(d)保持图像大小不变的同时以灰度级 128, 64 和 32 显示的图像；(e)~(h)以 16, 8, 4 和 2 灰度级显示的图像(原图像由 Vanderbilt 大学医学中心辐射学与放射学系的 David R. Pickens 博士提供)

83 ~ 85

例 2.2 和例 2.3 中的结果说明了分别改变 N 和 k 时对图像质量产生的影响。然而，这些结果也仅仅部分地回答了改变 N 和 k 如何影响图像的问题，因为我们还未考虑这两个参数之间可能存在的任何关系。Huang[1965]早期的研究试图通过实验来量化同时改变 N 和 k 时对图像质量所产生的影响。实验由一组主观测试组成，采用了类似于图 2.22 中所示的图像。妇女脸庞图像是包含较少细节的代表性图像，摄像师图像包含了中等程度的细节，人群图像相比之下包含有大量的细节。





a b c

图 2.22　(a)具有少量细节的图像；(b)具有中等程度细节的图像(图像由麻省理工学院提供)；(c)具有大量细节的图像

通过改变 N 和 k，生成了这三种类型的一组图像，然后要求观察者主观地按图像的质量对图像排序。最终结果以 Nk 平面中的等偏爱曲线形式汇总(图2.23 中显示了对应于图2.22 中图像的平均等偏爱曲线)。Nk 平面中的每一点表示一幅图像，该图像的 N 值和 k 值等于该点的坐标。位于等偏爱曲线上的点对应于有相等主观判定质量的图像。实验发现等偏爱曲线倾向于向右上方移动，但这三类图像的等偏爱曲线形状与图 2.23 中所示的那些曲线类似。这并不意外，因为曲线向右上方移动仅意味着更大的 N 值和 k 值，而这又意味着更好的图像质量。

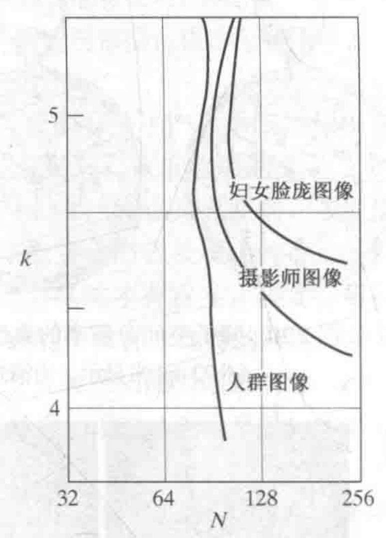


图 2.23　图 2.22 中三类图像的典型等偏爱曲线

当前讨论内容的关键是当图像中的细节增加时，等偏爱曲线会变得更加垂直。这一结果表明，对于有大量细节的图像，可能只需要较少的灰度级。例如，图 2.23 中对应于人群图像的等偏爱曲线接近垂直。这表明当 N 值固定时，这类图像的感觉质量与所用灰度级数(对于图2. 23所示的灰度级范围)近似无关。更有趣的是，还要注意，另两类图像在样本数增加的某些间隔内，感觉质量保持相同，但灰度级数实际上降低了。对于这一结果，最可能的原因是 k 的减小倾向于对比度的明显增加，人们通常感受到图像质量改善了的视觉效果。

2.4.4　图像内插

内插是在诸如放大、收缩、旋转和几何校正等任务中广泛应用的基本工具。本节的主要目标是介绍内插并用它调整图像的大小(收缩和放大)，这是基本的图像重取样方法。在诸如旋转和几何校正的应用中，内插的用法将在 2.6.5 节讨论。在第 4 章，我们还将回到这一主题，以详细地讨论图像重取样。

本质上，内插是用已知数据来估计未知位置的数值的处理。我们用一个简单的例子开始讨论这一话题。假设一幅大小为 500×500 像素的图像要放大 1.5 倍到 750×750 像素。一种简单的放大方法是创建一个假想的 750×750 网格，它与原始图像有相同的间隔，然后将其收缩，使它准确地与原

86 ~ 87

图像匹配。显然，收缩后的 750×750 网格的像素间隔要小于原图像的像素间隔。为了对覆盖的每个点赋以灰度值，我们在原图像中寻找最接近的像素，并把该像素的灰度赋给 750×750 网格中的新像素。完成对网格中覆盖的所有点的灰度赋值后，就把图像扩展到原来规定的大小，得到放大后的图像。

刚刚讨论的方法称为*最近邻内插*，因为这种方法把原图像中最近邻的灰度赋给了每个新位置（2.5 节将正式讨论邻近像素）。这种方法简单，但本节稍后将表明该方法有产生不希望的人为缺陷的倾向，如某些直边缘的严重失真。由于这一原因，实际上该方法并不常用。更实用的方法是*双线性内插*，在该方法中，我们用 4 个最近邻去估计给定位置的灰度。令 (x, y) 为我们想要赋以灰度值的位置（把它想象为前面描述的网格点）的坐标，并令 $v(x, y)$ 表示灰度值。对于双线性内插来说，赋值是由下面的公式得到的：

> 与该方法名称相反的是，双线性内插不是一种线性内插方法，因为其中包含有 xy 项。

$$v(x, y) = ax + by + cxy + d \tag{2.4-6}$$

式中，4 个系数可用由点(x, y)的 4 个最近邻点写出的未知方程确定。

正如我们马上将看到的那样，双线性内插给出了比最近邻内插好得多的结果，但随之而来的是计算量的增加。

下一种复杂度较高的方法是*双三次内插*，它包括 16 个最近邻点。赋予点 (x, y) 的灰度值是使用下式得到的：

$$v(x, y) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_{ij} x^i y^j \tag{2.4-7}$$

式中，16 个系数可用由点(x, y)的 16 个最近邻点写出的未知方程确定。观察式(2.4-7)，如果求和的上下限分别为 0 和 1，则其简化为式(2.4-6)的形式。通常，双三次内插在保持细节方面比双线性内插相对要好。双三次内插是商业图像编辑程序如 Adobe Photoshop 和 Corel Photopaint 的标准内插方法。

例 2.4　用于图像收缩和放大的内插方法的比较。

图 2.24(a)是与 2.20(d)相同的图像，它是由图 2.20(a)中分辨率为 1250 dpi 的图像先将分辨率降低到 72 dpi（大小由原来的 3692×2812 收缩到 213×162 像素），然后将缩小的图像再放大到原来的尺寸得到的。图 2.20(d)是采用最近邻内插法进行收缩与放大产生的。正如前面讨论的那样，图 2.24(a)的结果比较差。图 2.24(b)和图 2.24(c)是采用相同的过程，但分别采用双线性内插和双三次内插收缩和放大得到的。使用双线性内插得到的结果较最近邻内插有重大改进。双三次内插的结果较双线性内插的结果稍微清晰一些。图 2.24(d)与图 2.20(c)相同，它们都是由最近邻内插方法通过收缩与放大得到的。我们前面讨论的结论是分辨率降低到 150 dpi 时，图像开始显现退化。图 2.24(e)与图 2.24(f)分别显示了使用双线性内插和双三次内插收缩和放大图像的结果。尽管分辨率从 1250 dpi 降低到了 150 dpi，但与原图像相比，这两种内插方法再次显示了其强大的性能。与以前一样，双三次内插得到了稍微清晰一些的结果。

在内插中，有可能采用更多的邻点和更复杂的技术。例如，采用样条和小波，在某种情况下，可以得到比刚刚讨论的方法更好的结果。对于三维图形（Watt[1993]，Shirley[2002]）和医学图像处理（Lehmann et al.[1999]），保留精细细节特别重要，而对于普通数字图像处理，较少考虑额外的计算负担是合理的，所以双线性内插和双三次内插是人们选择的典型方法。

87
~
89

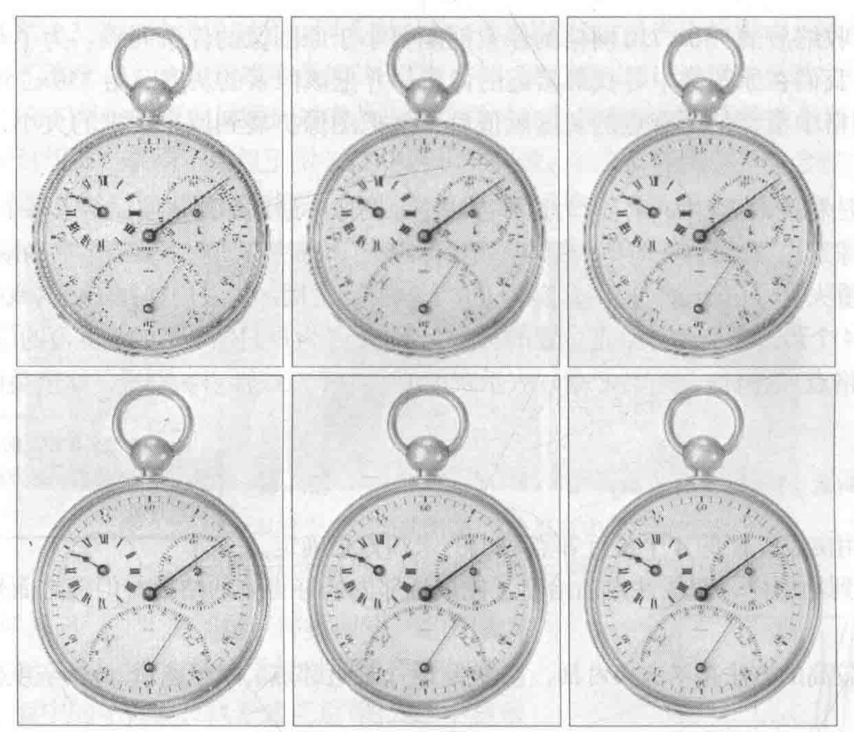


图 2.24 (a)降低到 72 dpi，并使用双线性内插方法将其放大到原大小(3692×2812 像素)的图像。该图像与图 2.20(d)相同；(b)使用双线性内插收缩和放大的图像；(c)与图像(b)相同但使用双三次内插的图像；(d)~(f)相同的图像序列，但它们是收缩到 150 dpi 而不是收缩到 72 dpi 的图像［图 2.24(d)与图 2.20(c)相同］。将图 2.24(e)和(f)与原图 2.20(a)比较，特别是后者

2.5 像素间的一些基本关系

本节介绍数字图像中像素间的几个重要关系。如前所述，图像由 $f(x,y)$ 表示。本节中引用某个特殊的像素时，我们通常使用小写字母，如 p 和 q。

2.5.1 相邻像素

位于坐标 (x,y) 处的像素 p 有 4 个水平和垂直的相邻像素，其坐标是

$$(x+1, y), (x-1, y), (x, y+1), (x, y-1)$$

这组像素称为 p 的 4 邻域，用 $N_4(p)$ 表示。每个像素距 (x,y) 一个单位距离，如果 (x,y) 位于图像的边界上，则 p 的某些相邻像素位于数字图像的外部。我们将在第 3 章处理这个问题。

p 的 4 个对角相邻像素的坐标是

$$(x+1, y+1), (x+1, y-1), (x-1, y+1), (x-1, y-1)$$

用 $N_D(p)$ 表示。这些点与 4 个邻点一起称为 p 的 8 邻域，用 $N_8(p)$表示。与前面一样，如果 (x,y) 位于图像的边界上，则 $N_D(p)$ 和 $N_8(p)$ 中的某些邻点会落入图像的外边。

2.5.2 邻接性、连通性、区域和边界


89 ~ 90

令 V 是用于定义邻接性的灰度值集合。在二值图像中，如果把具有 1 值的像素归诸于邻接像素，则 $V=\{1\}$。在灰度图像中，概念是一样的，但集合 V 一般包含有更多的元素。例如，灰度值范围为 0 到 255 的邻接像素中，集合 V 可能是这 256 个值的任何一个子集。考虑三种类型的邻接：

(a) 4邻接。如果 q 在集合 $N_4(p)$ 中，则具有 V 中数值的两个像素 p 和 q 是4邻接的。

(b) 8邻接。如果 q 在集合 $N_8(p)$ 中，则具有 V 中数值的两个像素 p 和 q 是8邻接的。

(c) m 邻接(混合邻接)。如果(i) q 在 $N_4(p)$ 中，或(ii) q 在 $N_D(p)$ 中，且集合 $N_4(p)\cap N_4(q)$ 中没有来自 V 中数值的像素，则具有 V 中数值的两个像素 p 和 q 是 m 邻接的。

> 我们分别使用符号 $\cap$ 和 $\cup$ 来表示集合的交与并。假设有集合 A 和 B，回忆可知它们的交集是既在集合 A 中又在集合 B 中的成员的元素集。这两个集合的并集是集合 A 的成员、集合 B 的成员或两者的成员的元素集。详细讨论见2.6.4节。

混合邻接是8邻接的改进。混合邻接的引入是为了消除采用8邻接时产生的二义性。例如，考虑图2.25(a)中对于 $V=\{1\}$ 的像素排列。位于图2.25(b)上部的3个像素显示了多重(二义性)8邻接，如虚线所示。这种二义性可以通过 m 邻接消除，如图2.25(c)所示。

从具有坐标 (x,y) 的像素 p 到具有坐标 (s,t) 的像素 q 的通路(或曲线)是特定的像素序列，其坐标为

$$(x_0,y_0),(x_1,y_1),\cdots,(x_n,y_n)$$

其中 $(x_0,y_0)=(x,y)$，$(x_n,y_n)=(s,t)$，且像素 (x_i,y_i) 和 (x_{i-1},y_{i-1}) 对于 $1\leqslant i\leqslant n$ 是邻接的。在这种情况下，n 是通路的长度。如果 $(x_0,y_0)=(x_n,y_n)$，则通路是闭合通路。可以依据特定的邻接类型定义4邻接、8邻接或 m 邻接。例如，如图2.25(b)所示，右上点和右下点之间的通路是8通路，而图2.25(c)中的通路是 m 通路。

图2.25　(a)像素的排列；(b) 8邻接像素(邻接性由虚线所示，注意二义性)；(c) m 邻接；(d)采用8邻接时，两个值为1的区域是邻接的；(e)如果在区域和背景间采用8邻接，则加圆圈的点是仅赋1值的像素的边界点的一部分；(f) 1值区域的内部边界不形成闭合通路，但其外部边界可以形成闭合通路

令 S 是图像中的一个像素子集。如果 S 的全部像素之间存在一个通路，则可以说两个像素 p 和 q 在 S 中是连通的。对于 S 中的任何像素 p，S 中连通到该像素的像素集称为 S 的连通分量。如果 S 仅有一个连通分量，则集合 S 称为连通集。

令 R 是图像中的一个像素子集。如果 R 是连通集，则称 R 为一个区域。两个区域，如果它们联合形成一个连通集，则区域 R_i 和 R_j 称为邻接区域。不邻接的区域称为不连接区域。在谈到区域时，我们考虑的是4邻接和8邻接。为使我们的定义有意义，必须指定邻接的类型。例如，如果仅使用8邻接，则图2.25(d)中的两个区域(由1组成的)是邻接的(根据前一段的定义，两个区域之间不存在4通路，它们的并集不是连通集)。

假设一幅图像包含有 K 个不连接的区域，即 $R_k, k=1,2,\cdots,K$，且它们都不接触图像的边界①。令 R_u 代表所有 K 个区域的并集，并且令 $(R_u)^c$ 代表其补集(回忆可知，集合 S 的补集是不在 S 中的点的集合)。我们称 R_u 中的所有点为图像的前景，而称 $(R_u)^c$ 中的所有点为图像的背景。

区域 R 的边界(也称为边缘或轮廓)是这样的点集，这些点与 R 的补集中的点邻近。换一种方式说，一个区域的边界是该区域中至少有一个背景邻点的像素集。这里再强调一下，我们必须指定用于定义邻接的连通性。例如，图2.25(e)中被圈出的点如果在区域及其背景间使用4连通，就不是1值区域边界的成员。基于这一规则，为处理这种情况，一个区域及其背景中的点之间的邻接要根据8连通来定义。

① 进行这种假设的目的在于避免处理特殊情形。这样做不会丧失一般性，因为如果一个或多个区域接触到图像的边界，我们可简单地使用1像素宽的背景值边界来填充图像。

90~92

前述定义有时称为区域的内边界，以便与其外边界相区分，外边界对应于背景边界。在开发追踪边界的算法时这一区别很重要。这种算法为了保证结果形成一个闭合通路，通常是沿外边界确立的。例如，在图2.25(f)中，1 值区域的内边界是该区域本身。这一边界并不满足先前给出的闭合通路的定义。另一方面，区域的外边界确实形成了一个围绕该区域的闭合通路。

如果 R 恰巧是整幅图像(假设这幅图像是像素的方形集合)，则边界由图像第一行、第一列和最后一行、最后一列的像素集合来定义。这个附加定义是需要的，因为一幅图像超过边界范围之外没有邻点。正常情况下，当我们提到一个区域时，指的是一幅图像的子集，并且区域边界中任何与图像边缘吻合的像素都作为区域边界的一部分全部包含于其中。

边缘的概念在涉及区域和边界的讨论中常常遇到。然而，这些概念之间有一个关键区别。一个有限区域的边界形成一条闭合通路，并且是"整体"概念。正如第 10 章详细讨论的那样，边缘由具有某些超过预先设定的阈值的导数值的像素形成。这样，边缘的概念就是基于在进行灰度级度量时不连续点的"局部"概念。把边缘点连接成边缘线段是可能的，并且有时以与边界对应的方法连接线段，但并不总是这样。边缘和边界吻合的一个例外是二值图像的情况。根据连通类型和所用的边缘算子(在第 10 章中讨论)，从二值区域提取边缘与区域边界是一样的，这很直观。第 10 章中我们会看到，概念上边缘考虑为灰度不连续，把边界考虑为闭合通路是有帮助的。

2.5.3 距离度量

对于坐标分别为 (x, y), (s, t) 和 (v, w) 的像素 p, q 和 z，如果

(a) $D(p, q) \geqslant 0$ [$D(p, q) = 0$ ，当且仅当 $p = q$]

(b) $D(p,q) = D(q,p)$ 且

(c) $D(p, z) \leqslant D(p, q) + D(q, z)$ 。

则 D 是距离函数或度量。p 和 q 间的欧几里得(欧氏)距离定义如下：

$$D_e(p, q) = \left[(x-s)^2 + (y-t)^2 \right]^{\frac{1}{2}} \tag{2.5-1}$$

对于距离度量，距点 (x, y) 的距离小于等于某个值 r 的像素，是中心在 (x, y) 且半径为 r 的圆平面。

p 和 q 间的距离 D_4 (又称为城市街区距离)由下式定义：

$$D_4(p, q) = |x-s| + |y-t| \tag{2.5-2}$$

在这种情况下，距 (x, y) 的距离 D_4 小于等于某个值 r 的像素形成一个中心在 (x, y) 的菱形。例如，距中心点 (x, y) 的距离 D_4 小于等于 2 的像素，形成固定距离的下列轮廓：

$$\begin{matrix} & & 2 & & \\ & 2 & 1 & 2 & \\ 2 & 1 & 0 & 1 & 2 \\ & 2 & 1 & 2 & \\ & & 2 & & \end{matrix}$$

其中 $D_4 = 1$ 的像素是 (x, y) 的 4 邻域。

p 和 q 间的 D_8 距离(又称为棋盘距离)由下式定义：

$$D_8(p, q) = \max(| x-s |,| y-t |) \tag{2.5-3}$$

92 ~ 93

在这种情况下，距 (x, y) 的 D_8 距离小于等于某个值 r 的像素形成中心在 (x, y) 的方形。例如，距中心点 (x, y) 的 D_8 距离小于等于 2 的像素形成下列固定距离的轮廓：

$$\begin{matrix} 2 & 2 & 2 & 2 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 \\ 2 & 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 2 & 2 \end{matrix}$$

其中 $D_8 = 1$ 的像素是 (x, y) 的 8 邻域。

注意，p 和 q 之间的 D_4 距离和 D_8 距离与任何通路无关，通路可能存在于各点之间，因为这些距离仅与该点的坐标有关。然而，如果选择考虑 m 邻接，则两点间的 D_m 距离用点间的最短通路定义。在这种情况下，两个像素间的距离将依赖于沿通路的像素值及其邻点值。例如，考虑如下排列的像素并假设 p, p_2 和 p_4 的值为 1，p_1 和 p_3 的值为 0 或 1：

$$\begin{matrix} & p_3 & p_4 \\ p_1 & p_2 & \\ p & & \end{matrix}$$

假设我们考虑值为 1 的像素邻接[即 $V = \{1\}$]。如果 p_1 和 p_3 是 0，则 p 和 p_4 间的最短 m 通路的长度（D_m 距离）是 2。如果 p_1 是 1，则 p_2 和 p 将不再是 m 邻接的（见 m 邻接的定义），并且最短 m 通路的长度变为 3（通路通过点 $p\ p_1\ p_2\ p_4$）。类似地，如果 p_3 是 1（且 p_1 是 0），此时最短的 m 通路距离也是 3。最后，如果 p_1 和 p_3 都为 1，则 p 和 p_4 间的最短 m 通路的长度为 4，此时通路通过点 $p\ p_1\ p_2\ p_3\ p_4$。

2.6 数字图像处理中所用数学工具的介绍

本节有两个主要目的：(1)介绍本书中所用的各种数学工具；(2)通过把它们用于各种基本图像处理任务中，帮助读者找到如何使用这些工具的“感觉”。后续讨论中会多次使用其中的一些工具。在后面的章节中，我们将根据需要来扩展这些工具的范围及应用。

2.6.1 阵列与矩阵操作

> 在继续阅读前，建议读者从本书网站上的 Tutorials 部分下载并学习复习材料。复习材料包括关于矩阵和向量、线性系统、集合论和概率的简介性知识。

包含一幅或多幅图像的阵列操作是逐个像素执行的。本章前面我们曾提到，图像可以等效地视为矩阵。事实上，在很多情况下，图像间的操作是用矩阵理论执行的（见2.6.6节）。基于这个原因，我们必须了解阵列与矩阵操作之间的区别。例如，考虑下面的 2×2 图像：

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{和} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

这两幅图像的阵列相乘是

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

另一方面，矩阵相乘由下式给出：

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

我们假设阵列操作贯穿全书，除非另做说明。例如，当我们谈到一幅图像的求幂时，意味着每个像素均进行求幂操作；当我们谈到一幅图像除以另一幅图像时，意味着在相应的像素对之间进行相除，等等。

93 ~ 95

2.6.2 线性操作与非线性操作

图像处理方法最重要的分类依据之一是，它是线性的还是非线性的。考虑一般的算子 H，该算子对于给定的输入图像 $f(x,y)$，产生一幅输出图像 $g(x,y)$：

$$H\left[f(x,y)\right]=g(x,y) \tag{2.6-1}$$

如果

$$H\left[a_i f_i(x,y)+a_j f_j(x,y)\right]=a_i H\left[f_i(x,y)\right]+a_j H\left[f_j(x,y)\right]=a_i g_i(x,y)+a_j g_j(x,y) \tag{2.6-2}$$

则称 H 是一个线性算子，其中 a_i, a_j，$f_i(x,y)$ 和 $f_j(x,y)$ 分别是任意常数和图像(大小相同)。式(2.6-2)指出输出是线性操作，因为两个输入的和，与分别对输入进行操作然后再求和得到的结果相同。另外，输入乘以常数的线性操作的输出，与乘以该常数的原始输入的操作的输出是相同的。第一个特性称为加性，第二个特性称为同质性。

作为一个简单的例子，假设 H 是求和算子 Σ，即该算子的功能是对输入简单地求和。为检验其线性，我们从式(2.6-2)的左侧开始，并试图证明它与右侧相等：

$$\begin{aligned}\sum\left[a_i f_i(x,y)+a_j f_j(x,y)\right]&=\sum a_i f_i(x,y)+\sum a_j f_j(x,y)\\&=a_i\sum f_i(x,y)+a_j\sum f_j(x,y)\\&=a_i g_i(x,y)+a_j g_j(x,y)\end{aligned}$$

这些是阵列求和，而不是图像的所有元素的求和。因此，单幅图像的求和是该图像本身。

其中第一步遵循求和是分布式的这样一个事实。因此，左边的展开等于式(2.6-2)的右边，从而我们得出该求和算子是线性的这一结论。

另一方面，考虑最大值操作，其功能是在图像中寻找像素的最大值。针对这一目的，证明该操作是非线性的最简单方法是，寻找一个测试式(2.6-2)时失败的例子。考虑下列两幅图像：

$$f_1=\begin{bmatrix}0 & 2\\2 & 3\end{bmatrix} \text{和}\ f_2=\begin{bmatrix}6 & 5\\4 & 7\end{bmatrix}$$

并假设令 $a_1=1$ 和 $a_2=-1$。为了对线性进行测试，我们再次从式(2.6-2)的左侧开始：

$$\max\left\{(1)\begin{bmatrix}0 & 2\\2 & 3\end{bmatrix}+(-1)\begin{bmatrix}6 & 5\\4 & 7\end{bmatrix}\right\}=\max\left\{\begin{bmatrix}-6 & -3\\-2 & -4\end{bmatrix}\right\}=-2$$

下一步，做右边，我们得到

$$(1)\max\left\{\begin{bmatrix}0 & 2\\2 & 3\end{bmatrix}\right\}+(-1)\max\left\{\begin{bmatrix}6 & 5\\4 & 7\end{bmatrix}\right\}=3+(-1)7=-4$$

此时，式(2.6-2)的左边和右边并不相等，因此我们证明了通常求最大值的操作是非线性的。

如后面三章所示，特别是第 4 章和第 5 章，线性操作特别重要，因为它们是以可用于图像处理的大量理论和实践结果为基础的。非线性系统很难被很好地理解，因此其应用范围受到了更多的限制。但在后面几章中，我们会碰到一些性能远优于线性操作的非线性图像处理操作。

2.6.3 算术操作

如2.6.1节中讨论的那样，图像间的算术操作是阵列操作，即算术操作在相应的像素对之间执行。4 种算术操作表示为

95 ~ 96

$$\begin{aligned} s(x, y) &= f(x, y) + g(x, y) \\ d(x, y) &= f(x, y) - g(x, y) \\ p(x, y) &= f(x, y) \times g(x, y) \\ v(x, y) &= f(x, y) \div g(x, y) \end{aligned} \tag{2.6-3}$$

它可理解为是在f和g中相应的像素对之间执行操作，其中$x = 0, 1, 2, \cdots, M-1$，$y = 0, 1, 2, \cdots, N-1$。通常，M和N是图像的行和列。很明显，s, d, p和v是大小为$M \times N$的图像。注意，按照刚才定义的方式，图像算术操作涉及相同大小的图像。下面的几个例子指出在数字图像处理中，算术操作起着重要的作用。

例2.5　针对降噪的带噪图像相加（平均）。

令$g(x, y)$是无噪声图像$f(x, y)$被加性噪声$\eta(x, y)$污染后的图像，即

$$g(x, y) = f(x, y) + \eta(x, y) \tag{2.6-4}$$

这里假设在每一对坐标(x, y)处，噪声是不相关的[①]，并且其均值为零。以下步骤的目的是通过一组带噪图像$\{g_i(x, y)\}$的相加减少噪声。

如果噪声满足刚才声明的约束，那么可以证明（见习题2.20），如果图像$\bar{g}(x, y)$是通过对K幅不同的噪声图像进行平均形成的：

$$\bar{g}(x, y) = \frac{1}{K}\sum_{i=1}^{K} g_i(x, y) \tag{2.6-5}$$

且遵循

$$E\{\bar{g}(x, y)\} = f(x, y) \tag{2.6-6}$$

和

$$\sigma^2_{\bar{g}(x, y)} = \frac{1}{K}\sigma^2_{\eta(x, y)} \tag{2.6-7}$$

式中$E\{\bar{g}(x, y)\}$是$\bar{g}$的期望值，$\sigma^2_{\bar{g}(x, y)}$和$\sigma^2_{\eta(x, y)}$分别是$\bar{g}$和η在所有坐标(x, y)处的方差，那么在平均图像中的任意一点处的标准差（方差的均方根）是

$$\sigma_{\bar{g}(x, y)} = \frac{1}{\sqrt{K}}\sigma_{\eta(x, y)} \tag{2.6-8}$$

随着K的增大，式（2.6-7）和式（2.6-8）指出，在每个位置(x, y)处的像素值的变化（就像方差或标准差度量的那样）将减小。因为$E\{\bar{g}(x, y)\} = f(x, y)$，这意味着在求平均过程中所用的带噪图像的数量增加时，$\bar{g}(x, y)$将逼近$f(x, y)$。实际上，为了避免输出图像带来模糊和其他人为缺陷，图像$g_i(x, y)$必须已配准（对齐）。

图像平均的一种重要应用是在天文学领域，在该领域，由于在非常低的照度下成像常常会导致传感器噪声，以至于单幅图像无法分析。图2.26(a)显示了一幅8比特图像，其中加入了均值为0、标准差为64个灰度级的高斯噪声模拟退化。这些典型的低照度条件下获取的图像，对所有实用目的来说都不可用。图2.26(b)～(f)显示了分别对5, 10, 20, 50和100幅图

> 该例中显示的图像来自称为NGC 3314的星系对，由NASA的哈勃空间望远镜拍摄。NGC 3314离地球约1.4亿光年，位于南半球长蛇座方向。最亮的几颗星在由星尘组成的前星系的中心处形成了风车形状。

① 回忆可知，均值为m的随机变量z的方差定义为$E[(z-m)^2]$，其中$E\{\cdot\}$是该参量的期望值。两个随机变量z_i和z_j的协方差定义为$E[(z_i - m_i)(z_j - m_j)]$。如果变量是不相关的，则它们的协方差为零。

像平均的结果。我们看到，图2.26(e)，即由 $K=50$ 得到的结果已经相当清晰。图 2.26(f)的图像是由 100 幅带噪图像平均得到的结果，与图2.26(e)比较仅稍有改进。

相加是连续积分的离散形式。在天文观测中，一种与刚刚描述的方法等同的处理是使用 CCD（见 2.3.3 节）或类似传感器的累积能力，通过长时间地观察同一场景来达到降噪的目的。冷却也常常用于降低传感器噪声。然而，最终结果类似于对一组噪声数字图像进行平均操作。

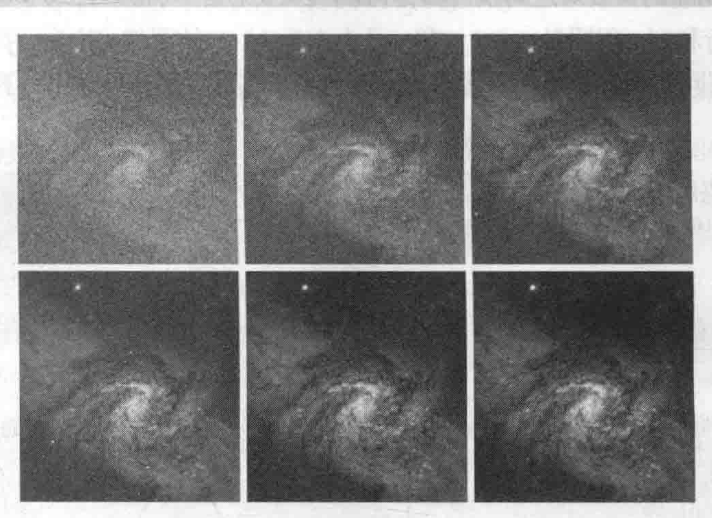


图 2.26 (a)被加性高斯噪声污染的星系对 NGC 3314 图像；(b)~(f)分别是对 5, 10, 20, 50, 100 幅噪声图像平均的结果(原图像由 NASA 提供)

例 2.6 增强差别的图像相减。

图像相减经常用于增强图像之间的差别。例如，图 2.27(b)中的图像是通过把图 2.27(a)中的每个像素的最低有效比特置 0 得到的。在视觉上，这些图像都不可辨别。然而，如图2.27(c)所示，从一幅图像减去另一幅图像，则清楚地显示了它们的差别。在差值图像中，黑(0)值指出了在该位置图 2.27(a)和图 2.27(b)之间没有差别。

作为另一个例子，我们简单地讨论称之为模板模式放射成像的医学成像领域，这是图像相减在商业上成功且高受益的应用。考虑如下形式的图像差：

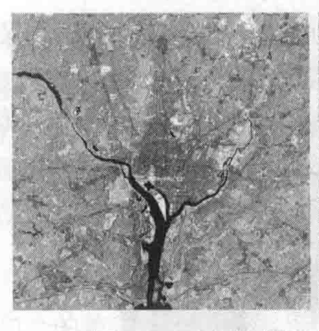
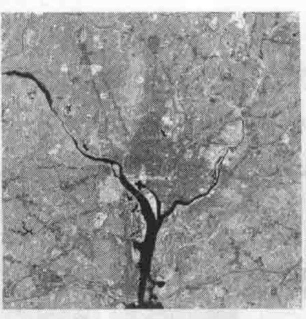
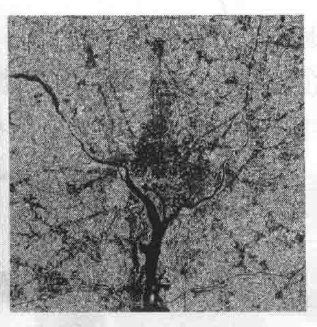
$$g(x, y) = f(x, y) - h(x, y) \tag{2.6-9}$$

> 通过图像相减来更改检测也用于图像分割，图像分割是第 10 章的主题。

在这种情况下，模板 $h(x, y)$ 是病人身体一个区域的 X 射线图像，该图像由放在 X 射线源对面的电视摄像机(代替传统的 X 射线胶片)获取。过程由以下几步组成：将一种 X 射线造影剂注入病人的血流，摄取一系列与 $h(x, y)$解剖区域相同的活体图像［它们的样品表示为 $f(x, y)$］，从一系列注射过造影剂的活体图像中减去模板 $h(x, y)$。从每幅活体图像中减去模板的最终结果是这样的区域，即出现在输出图像 $g(x, y)$ 中用以区分 $f(x, y)$ 和 $h(x, y)$ 之间差别的区域，目的是增强细节。因为图像可以按电视速率获取，这一过程基本上给出了一段显示造影剂怎样在被观察区域通过动脉传播的影片。

图 2.28(a)显示了在碘介质注射到血流中前，病人头部上方的一幅模板的 X 射线图像，图 2.28(b)是在注射碘介质后得到的一幅活体图像。图2.28(c)是图2.8(a)与图2.8(b)的差。在这幅图像中，一些较细的血管结构清晰可见。图 2.28(d)的差别很清楚，它是由(c)增强了对比度后得到的(下一章中将讨论对比度增强)。图 2.28(d)是介质在大脑内通过血管传播的一幅清晰“图形”。

98
~
99

a b c

图 2.27 (a)华盛顿特区的红外图像；(b)将图像(a)中的每个最低阶比特置 0 得到的图像；(c)两幅图像的差，为清楚起见，图像已标定到区间[0, 255]

a b c d

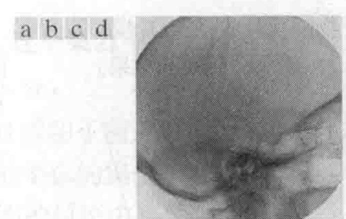





图 2.28 数字减影血管造影：(a)模板图像；(b)活体图像；(c)图(a)和图(b)的差值图像；(d)增强后的差值图像［图(a)和图(b)由荷兰 Utrecht 大学图像科学研究所医学中心提供］

例 2.7 使用图像相乘和相除来校正阴影。

图像相乘(或相除)的一种重要应用是阴影校正。假设一个成像传感器产生可由 $f(x, y)$ 表示的完美图像与阴影函数 $h(x, y)$ 的乘积来建模的图像，即 $g(x, y) = f(x, y)h(x, y)$。如果 $h(x, y)$ 已知，那么我们可以使用 $h(x, y)$ 的反函数［即 $g(x, y)/h(x, y)$］乘以感知图像的方法得到 $f(x, y)$。如果 $h(x, y)$ 未知，但图像系统可以访问，我们可以通过对具有恒定灰度的目标成像得到一个近似的阴影函数。传感器不可用时，我们常常如 9.6 节讨论的那样直接从图像估计阴影模式。图2.29 显示了阴影校正的一个例子。

图像相乘的另一种普通应用是模板操作，也称为感兴趣区域（ROI）操作。图2.30 中说明的处理仅将模板图像与一幅给定的图像相乘，模板图像的 ROI 区域为 1，其他区域为 0。模板图像中的 ROI 可能不止一个，虽然由于实现容易而常常使用矩形形状，但 ROI 的形状可以是任意的。





a b c

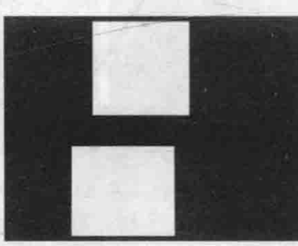
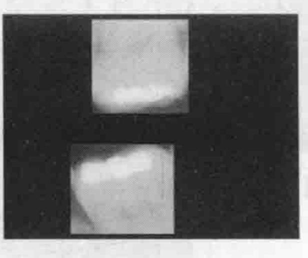
图 2.29 阴影校正：(a)钨丝及其支撑物的阴影 SEM 图像，该图像放大了近 130 倍；(b)阴影模式；(c)图(b)的倒数与图(a)相乘的结果(原图像由俄勒冈尤金大学地质科学系的 Michael Shaffer 提供)

在结束这一节的讨论之前，对图像算术操作的实现做些评述是有益的。实际上，大多数图像使用 8 比特显示(即使 24 比特的彩色图像也是由分立的三个 8 比特通道组成的)。这样，我们认为图像的灰度值范围为 0 ~255。当图像以标准格式存储时，如 TIFF 或 JPEG，图像的灰度值将自动转换到该范围。然而，转换所用的方法取决于所用的系统。例如，两幅 8 比特图像的不同取值范围可以是从最小的–255 到最大的+255，图像之和的值可能在范围 0 到 510 内。许多软件包在把图像转换为 8 比特

99 ~ 101

图像时，仅简单地把所有的负值转换为 0，而把超过 255 的值转换为 255。给定一幅图像f，保证图像间算术操作的整个值域落入某个固定比特数的方法如下。首先，我们执行操作

$$f_m = f - \min(f) \tag{2.6-10}$$

a b c

图 2.30 (a) 牙齿的 X 射线数字图像；(b) 使用填充物隔离牙齿的 ROI 模板(白色对应于 1，黑色对应于 0)；(c) 图(a)和图(b)相乘的结果

该操作生成最小值为 0 的一幅图像。然后，执行操作

$$f_s = K\left[f_m / \max(f_m)\right] \tag{2.6-11}$$

该操作生成一幅标定的图像f_s，其值在[0, K]范围内。在处理 8 比特图像时，置$K = 255$，我们得到一幅灰度范围从 0 到 255 的 8 比特满标度图像。类似的方法可用于 16 比特或更高比特的图像。这种方法可用于所有的算术操作。在执行除法操作时，我们还有额外的要求，即需要将一个较小的数加到除数图像的像素上，以避免除以 0 的情况。

2.6.4 集合和逻辑操作

本节简单介绍一些重要的集合和逻辑操作，以及模糊集合的概念。

基本集合操作

令A为一个实数序对组成的集合。如果$a = (a_1, a_2)$是A的一个元素，则将其写成

$$a \in A \tag{2.6-12}$$

同样，如果a不是A的一个元素，则写成

$$a \notin A \tag{2.6-13}$$

不包含任何元素的集合称为空集，用符号$\varnothing$表示。

集合由两个大括号中的内容表示，即$\{\cdot\}$。例如，当我们将一个表达式写成$C = \{w \mid w = -d, d \in D\}$的形式时，所表达的意思是：集合$C$是元素$w$的集合，而$w$是通过用−1 与集合$D$中的所有元素相乘得到的。该集合用于图像处理的一种方法是，令集合的元素为图像中表示区域(物体)的像素的坐标(整数序对)。

如果集合A中的每个元素又是另一个集合B中的一个元素，则称A为B的子集，表示为

$$A \subseteq B \tag{2.6-14}$$

两个集合A和B的并集表示为

$$C = A \cup B \tag{2.6-15}$$

这个集合包含集合A和B中的所有元素。类似地，两个集合A和B的交集表示为

$$D = A \cap B \tag{2.6-16}$$

101 ~ 102

这个集合包含的元素同时属于集合A和B。如果A和B两个集合没有共同的元素，则称这两个集合是不相容的或互斥的。此时，

$$A \cap B = \varnothing \tag{2.6-17}$$

全集 U 是给定应用中的所有元素的集合。根据这一定义，给定应用的所有集合元素是对于该应用定义的全部成员。例如，如果处理实数集合，则集合的全集是实数域，它包含所有的实数。在图像处理中，我们一般将全集定义为包含图像中所有像素的正方形。

集合 A 的补集是不包含于集合 A 的元素所组成的集合，表示为

$$A^c = \{w \mid w \notin A\} \tag{2.6-18}$$

集合 A 和 B 的差表示为 $A-B$，定义为

$$A - B = \{w \mid w \in A, w \notin B\} = A \cap B^c \tag{2.6-19}$$

可以看出，这个集合中的元素属于 A 而不属于 B。例如，我们可以根据 U 并做集合的差操作来定义 A^c，即 $A^c = U - A$。

图2.31说明了前述概念，全集是包含在所示正方形中的坐标的集合，并且集合 A 和 B 是包含在所示边界内的坐标的集合。在每一幅图中，集合操作的结果用灰色显示[①]。

在前面的讨论中，集合成员是以位置（坐标）为基础的。处理图像时，一个隐含的假设是，集合内所有像素的灰度相同，因为我们还没有定义关于灰度值的集合操作（即我们还未指定两个集合的交集中的灰度是什么）。图2.31所示操作的唯一方法可以这样确定，如果包含这两个集合的图像是二值的，在这种情况下，我们可以说集合成员是基于坐标的，并假设这两个集合的所有成员具有相同的灰度。我们将在下一节做更详细的讨论。

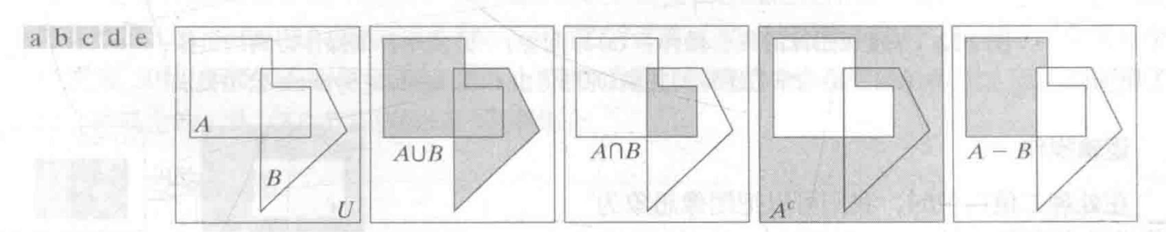


图 2.31　(a) 二维空间中的两个坐标集合 A 和 B；(b) A 和 B 的并集；(c) A 和 B 的交集；(d) A 的补集；(e) A 和 B 的差。在 (b) 到 (e) 中，阴影区域表示指定集合操作的成员

在处理灰度图像时，前述概念不再适用，因为我们必须指定来自集合操作结果的所有像素的灰度。事实上，正如我们将在 3.8 节和 9.6 节中看到的那样，灰度值的并集操作和交集操作通常分别定义为相应像素对的最大和最小，而补集操作定义为常数与图像中每个像素的灰度间的两两之差。我们处理相应像素对的事实告诉我们，灰度级集合操作是如2.6.1节定义的阵列操作。下面的例子是涉及灰度级图像的集合操作的简单说明。我们将进一步讨论上面两节提到的概念。

例 2.8　图像灰度的集合操作。

令灰度级图像的元素用集合 A 来表示，如 2.4.2 节提到的那样，这些元素是三元组形式 (x, y, z)，其中 x 和 y 是空间坐标，z 是灰度。我们可以将 A 的补集定义为 $A^c = \{(x, y, K - z) \mid (x, y, z) \in A\}$，它简单地表明像素集合 A 中的灰度已从常数 K 中减去。该常数等于 $2^k - 1$，其中 k 是用于表示 z 的灰度的比特数。令 A 表示图 2.32(a) 中的 8 比特灰度图像，并假设我们想要用集合操作形成一个负 A。简单地形成集合 $A_n = A^c = \{(x, y, 255 - z) \mid (x, y, z) \in A\}$。注意，坐标不变，因此 A_n 是与 A 大小相同的图像。图 2.32(b) 示出了这一结果。

① 式 (2.6-12) 至式 (2.6-19) 中的操作是代数集的基础，它从 $A \cup B = B \cup A$ 和 $A \cap B = B \cap A$ 这样的交换律开始，形成了集合操作的广泛理论基础。尽管代数集的处理超出了当前讨论的范围，但读者应了解这一点。

两个灰度集合 A 和 B 的并集可定义为集合

$$A \bigcup B = \{\max_{z}(a,b) \mid a \in A, b \in B\}$$

也就是说，两个灰度集合(图像)的并集是一个由空间相应元素对间的最大灰度形成的阵列。注意坐标同样不变，因此 A 和 B 的并集是大小与这两幅图像相同的图像。正如说明的那样，假设 A 仍表示图2.32(a)中的图像，并且令 B 表示与 A 大小相同的方形阵列，但其中所有的 z 值等于 A 中元素的平均灰度 m 的 3 倍。图2.32(c)显示了执行这种并集操作的结果，其中所有超过 $3m$ 的值呈现为 A 的值，而所有其他像素的值为 $3m$，即中间灰度值。

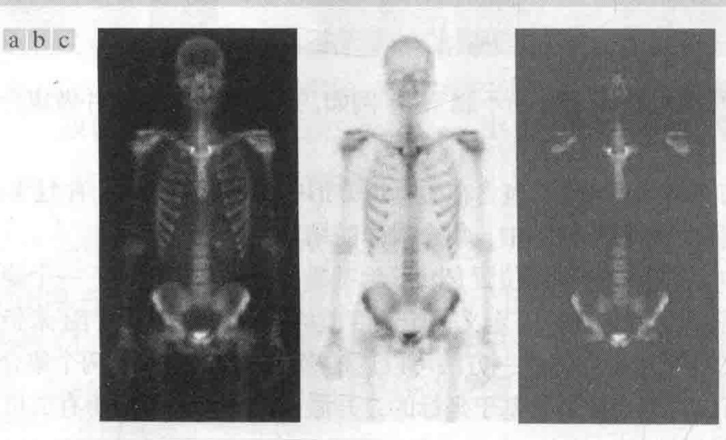


图 2.32　灰度级图像的集合操作：(a)原图像；(b)使用补集操作得到的负像；(c)图(a)与常数图像的并集(原图像由 G. E. Medical Systems 公司提供)

逻辑操作

在处理二值图像时，我们可以把图像想象为像素集合的前景(1 值)与背景(0 值)。然后，如果我们将区域(目标)定义为由前景像素组成，则图 2.31 中说明的集合操作就变成了二值图像中目标坐标间的操作。处理二值图像时，OR、AND 和 NOT 逻辑操作就是指普通的并、交和求补操作，其中“逻辑”一词来自逻辑理论，在逻辑理论中，1 代表真，0 代表假。

考虑由前景像素组成的区域(集合)A 和 B。这两个集合的 OR(或)操作结果不是属于 A，就是属于 B，或者属于两者。AND 操作是共同属于 A 和 B 的元素的集合。集合 A 的 NOT 操作是不在 A 中的元素的集合。因为我们要处理图像，如果 A 是给定的前景像素的集合，那么 NOT(A)是图像中不在 A 中的所有像素的集合，这些像素是背景像素，并有可能是其他前景像素。我们可以将该操作想象为：把 A 中的所有像素转换为 0(黑色)，并把所有不在 A 中的元素转换为 1(白色)。图 2.33

105~106

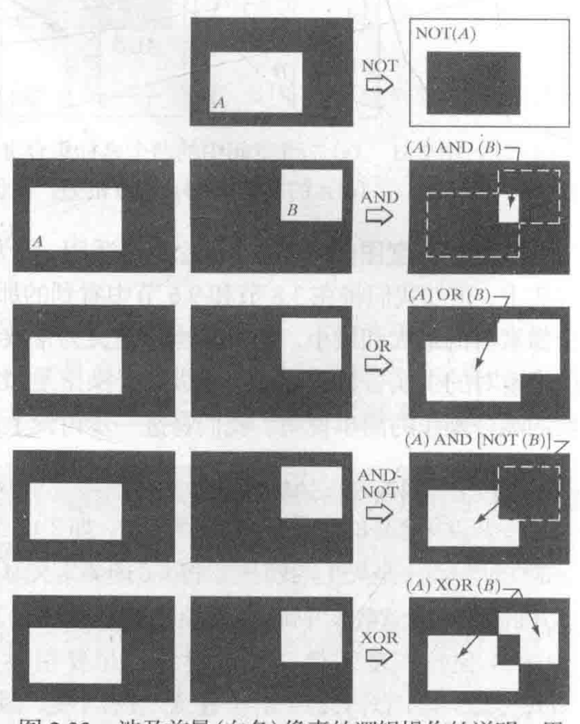


图 2.33　涉及前景(白色)像素的逻辑操作的说明。黑色代表二进制数 0，白色代表二进制数 1。虚线不是结果的一部分，仅作为参考显示

说明了这些操作。注意，显示该操作结果的第 4 行是前景像素集合，它属于 A 但不属于 B，这就是式(2.6-19)中集合差的定义。图中的最后一行是 XOR(异或)操作的结果，该结果是属于 A 或 B 的前景像素的集合，但不是两者的前景像素的集合。可以说，前述操作是区域间的操作，很明显，区域可以是不规则的和大小不同的。这与前面讨论的灰度级操作矛盾，前面的灰度操作是阵列操作并要求空间维数相同。也就是说，灰度集合操作涉及整个图像，而不是图像的区域。

我们仅在实现 AND，OR 和 NOT 逻辑算子时需要涉及理论知识，因为这三个算子是功能完备的。换句话说，任何其他逻辑操作都可以仅用这三个基本函数执行，如图 2.33 中的第 4 行所示，这里我们用 AND 和 NOT 实现了差集操作。逻辑操作广泛用于图像形态学处理，这是第 9 章的主题。

模糊集合

前述的集合和逻辑结果很“干脆”，在这种意义下，元素要么是要么不是集合中的成员。这在某些应用中会受到严重限制。考虑一个简单的例子。假如我们希望把世界上的所有人分为年轻人和非年轻人。使用“干脆”集合，令 U 代表所有人，令 A 是 U 的子集，我们把它称为年轻人集合。为了形成集合 A，我们需要一个隶属度函数，该函数可以对 U 中的每个元素(人)赋 1 值或 0 值。如果 U 中的元素赋值为 1，则该元素就是 A 的一个成员，否则就不是 A 的一个成员。因为我们是在处理二值逻辑，隶属度函数简单地定义为一个阈值，低于该阈值的人考虑为年轻人，高于该阈值的人考虑为非年轻人。假如我们定义任何 20 岁或更年轻的人是年轻人。我们马上面临的困难是：年龄为 20 岁零 1 秒的人将不是年轻人集合的成员。这一限制出现在我们用于分类年轻人年龄的阈值上。什么是我们意味的“年轻”需要更多的灵活性，即需要从年轻到非年轻逐步过渡。模糊集合理论使用隶属度函数来实现这种概念，该函数在数值 1(定义为年轻)和 0(定义为非年轻)之间逐步过渡。使用模糊集合，我们可以声明一个人的年轻度为 50%(年轻和非年轻过渡的中间)。换句话说，年龄是一个不精确的概念，而模糊逻辑了提供处理这种概念的工具。3.8 节将详细探讨模糊集合。

2.6.5　空间操作

空间操作直接在给定图像的像素上执行。我们把空间操作分为三大类：(1)单像素操作；(2)邻域操作；(3)几何空间变换。

单像素操作

我们对数字图像执行的最简单的操作是，以灰度为基础改变单个像素的值。这类处理可以用一个形如下式的变换函数 T 来描述：

$$s = T(z) \tag{2.6-20}$$

式中，z 是原图像中像素的灰度，s 是处理后的图像中相应像素的(映射)灰度。例如，图2.34示出了得到一幅 8 比特负图像的变换，图2.32(b)中的图像是利用集合操作得到的。我们将在第3章中讨论指定灰度变换功能的一些技术。

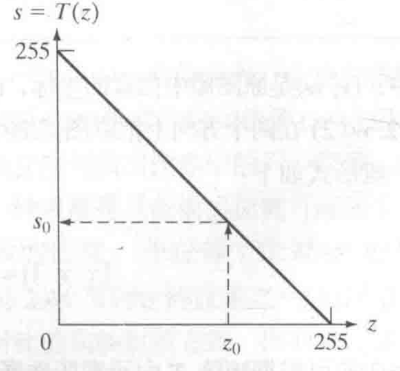


图 2.34　用于得到 8 比特图像的负图像的灰度变换函数。虚线箭头显示了任意输入灰度值 z_0 到对应输出值 s_0 的变换

邻域操作

令 S_{xy} 代表图像 f 中以任意一点 (x, y) 为中心的一个邻域的坐标集。邻域处理在输出图像 g 中的相同坐标处生成一个相应的像素，该像素的值由输入图像中坐标在 S_{xy} 内的像素经指定操作决定。例如，

106 ~ 107

假设指定的操作是计算在大小为$m \times n$、中心在(x, y)的矩形邻域中的像素的平均值。这个区域中像素的位置组成集合S_{xy}。图2.35(a)和(b)说明了这一过程。我们可以用公式的形式将这一操作描述为

$$g(x,y)=\frac{1}{mn}\sum_{(r,c)\in S_{xy}} f(r,c) \tag{2.6-21}$$

式中r和c是像素的行和列坐标，这些坐标是S_{xy}中的成员。图像g是这样得到的：改变坐标(x,y)，以便邻域的中心在图像f中从一个像素到另一个像素移动，并在每个新位置重复邻域操作。例如，图2.35(d)中的图像是用大小为41×41的邻域按这种方式创建的。最终结果是在原图像中执行局部模糊。例如，这种类型的处理可消除小细节，并在图像中相应的大区域实施"斑点"补偿。我们将在第3章、第5章和本书的其他几处地方讨论邻域处理。

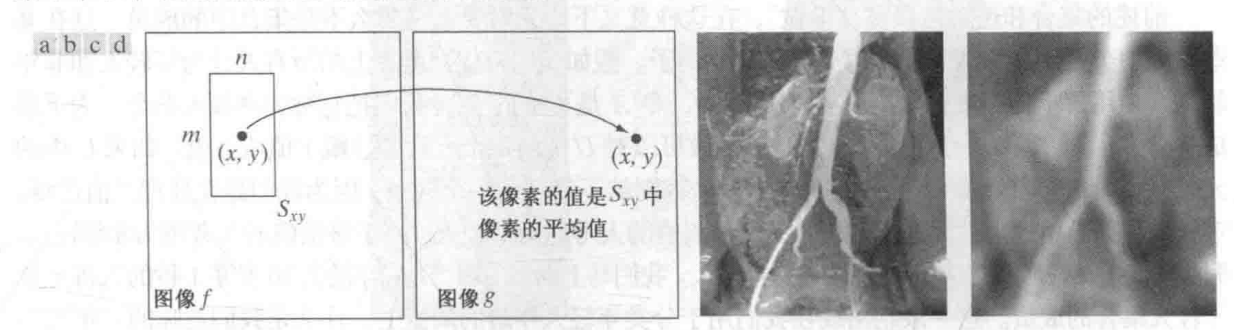


图2.35 使用邻域处理的局部平均：(a)～(b)说明了矩形邻域的过程；(c)1.3.2节讨论过的大动脉血管造影；(d)将$m=n=41$代入式(2.6-21)后的结果。图像大小为790×686像素

几何空间变换和图像配准

几何变换改进图像中像素间的空间关系。这些变换通常称为橡皮膜变换，因为它们可视为在一块橡皮膜上印刷一幅图像，然后根据预定的一组规则拉伸该橡皮膜。在数字图像处理中，几何变换由两个基本操作组成：(1)坐标的空间变换；(2)灰度内插，即对空间变换后的像素赋灰度值。

坐标变换可由下式表示：

$$(x,y)=T\{(v,w)\} \tag{2.6-22}$$

式中，(v, w)是原图像中像素的坐标，(x, y)是变换后图像中像素的坐标。例如，变换$(x,y)=T\{(v,w)\}=(v/2, w/2)$在两个方向上把原图像缩小一半。最常用的空间坐标变换之一是仿射变换(Wolberg[1990])，其一般形式如下：

$$[x\ \ y\ \ 1]=[v\ \ w\ \ 1]\boldsymbol{T}=[v\ \ w\ \ 1]\begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix} \tag{2.6-23}$$

这个变换可根据矩阵$\boldsymbol{T}$中元素所选择的值，对一组坐标点做尺度、旋转、平移或偏移。表2.2说明了用于实现这些变换的矩阵值。式(2.6-23)中矩阵表示的真实作用提供了把一系列操作连接在一起的框架。例如，如果我们想要调整一幅图像的大小，对其旋转，并把处理结果移动到某一位置，那么可以简单地构成一个3×3矩阵，然后乘以表2.2中的尺度、旋转和平移矩阵。

前面的变换把一幅图像上的像素重新定位到一个新位置。为了完成该处理，我们还必须对这些新位置赋灰度值。该任务可用灰度内插的方法完成。灰度内插这一主题已在2.4.4节讨论过。在那一节开始，我们用一个放大图像的例子讨论了给新像素位置赋灰度值的问题。正像表2.2中第二行详细表

108 ~ 109

述的那样，放大是简单的尺度变换，且其分析类似于我们对图像放大那样，可用于从表 2.2 中的其他变换得到的对重定位像素赋灰度值的问题。如 2.44 节那样，在做这些变换时，我们可考虑最近邻、双线性和双三次内插技术。

表 2.2　基于式(2.6-23)的仿射变换

变换名称	仿射矩阵 T	坐标公式	例子
恒等变换	$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$	$x=v$ $y=w$	
尺度变换	$\begin{bmatrix}c_x&0&0\\0&c_y&0\\0&0&1\end{bmatrix}$	$x=c_xv$ $y=c_yw$	
旋转变换	$\begin{bmatrix}\cos\theta&\sin\theta&0\\-\sin\theta&\cos\theta&0\\0&0&1\end{bmatrix}$	$x=v\cos\theta-w\sin\theta$ $y=v\sin\theta+w\cos\theta$	
平移变换	$\begin{bmatrix}1&0&0\\0&1&0\\t_x&t_y&1\end{bmatrix}$	$x=v+t_x$ $y=w+t_y$	
(垂直)偏移变换	$\begin{bmatrix}1&0&0\\s_v&1&0\\0&0&1\end{bmatrix}$	$x=vs_v+w$ $y=w$	
(水平)偏移变换	$\begin{bmatrix}1&s_h&0\\0&1&0\\0&0&1\end{bmatrix}$	$x=v$ $y=s_hv+w$	

实际上，我们可以按两种基本方法来使用式(2.6-23)。第一种方法称为前向映射，它由扫描输入图像的像素，并在每个位置(v, w)用式(2.6-23)直接计算输出图像中相应像素的空间位置(x, y)组成。前向映射算法的一个问题是，输入图像中的两个或多个像素可被变换到输出图像中的同一位置，这就产生了如何把多个输出值合并到一个输出像素的问题。另外，一种可能是某些输出位置可能完全没有要赋值的像素。第二种方法，称为反向映射，扫描输出像素的位置，并在每个位置(x, y)使用$(v, w)=\boldsymbol{T}^{-1}(x, y)$计算输入图像中的相应位置。然后，内插(使用 2.44 节讨论的技术之一)使用最近的输入像素之一决定输出像素的灰度值。对于实现来说，反向映射比前向映射更有效，因而被许多空间变换的商业实现所采用(如 MATLAB 就采用该方法)。

例 2.9　图像旋转与灰度内插。

本例的目的是用仿射变换说明图像旋转。图 2.36(a)显示了一幅 300 dpi 的图像，图 2.36(b)~(d)是将原图像旋转 21° 并分别使用最近邻、双线性和双三次内插操作后的结果。旋转是在保持直线特性方面最苛求的几何变换之一。如我们在图中看到的那样，最近邻内插产生了最大的锯齿边缘，如 2.4.4 节那样，双线性内插得到了明显改进的结果。如以前那样，使用双三次内插产生了稍微清晰一些的结果。事实上，

109 ~ 111

如果比较图 2.36(c)和图 2.36(d)中放大了的细节，就会在子图像的中间注意到由亮到暗变换引入的大量垂直灰度块，而图 2.36(c)中的垂直灰度块数量要比图 2.36(d)中的相应垂直灰度块数量多，这说明字母的边缘更清晰。类似的结果可用表 2.2 中的其他空间变换，按要求的内插得到(恒等变换不会得到类似的结果，增量为整数个像素的平移变换也不会得到类似的结果)。这个例子是使用前一段讨论的反向映射方法实现的。

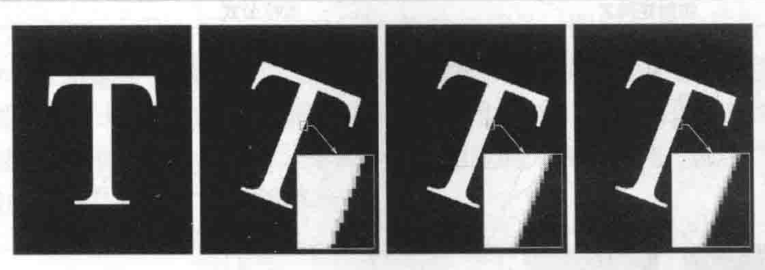


图 2.36　(a)字母 T 的 300 dpi 图像；(b)旋转 21° 并用最近邻内插给经空间变换后的像素赋灰度值后的图像；(c)旋转 21° 并用双线性内插赋灰度值后的图像；(d)旋转 21° 并用双三次内插赋灰度值后的图像。放大部分显示了三种内插方法的边缘细节

图像配准是数字图像处理的一种重要应用，用于对齐两幅或多幅相同场景的图像。在前述讨论中，我们已知为达到希望的几何变换的变换函数形式。在图像配准中，我们有可用的输入图像和输出图像，但从输入图像产生输出图像的特定变换通常是不知道的。这样，问题就是估计变换函数，然后用它配准两幅图像。为搞清楚该术语，输入图像是我们希望变换的图像，而所谓的参考图像是我们想要配准输入图像的图像。

例如，感兴趣的是对齐(配准)在近似相同时间但使用不同成像系统获取的两幅或多幅图像，如一幅是用 MRI(核磁共振成像)扫描器获取的图像，另一幅是用 PET(正电子断层)扫描器获取的图像；或在不同的时间使用相同的设备获取的图像，如给定位置间隔几天、几个月甚至几年的卫星图像。无论哪种情况，合并这些图像或执行定量分析，以及做比较，都要求对由观察视角、距离、方向、传感器分辨率、目标位置移动及其他因素导致的几何畸变进行修正。

解决刚才讨论的这个问题的主要方法之一是使用约束点(也称为*控制点*)，这些点是在输入图像和参考图像中其位置恰好已知的相应点。有很多选择约束点的方法，可以交互式地选择这些点，也可以使用试图自动检测这些点的算法。在某些应用中，成像系统在成像传感器中会嵌入有形的人为物质（如小金属物)。这会在系统获取的图像上直接产生一系列*已知点*(称为*网状标记*)，这些点的作用是指导如何建立约束点。

估计变换函数问题是建模问题之一。例如，假设在输入图像和参考图像中有一组 4 个约束点。基于双线性近似的简单模型由如下两式给出：

$$x = c_1 v + c_2 w + c_3 vw + c_4 \tag{2.6-24}$$

和

$$y = c_5 v + c_6 w + c_7 vw + c_8 \tag{2.6-25}$$

其中，在估计阶段，(v, w) 和 (x, y) 分别是输入图像和参考图像中约束点的坐标。如果在两幅图像中我们有 4 对相应的约束点，则可用式(2.6-24)和式(2.6-25)写出 8 个方程，然后用它们解出 8 个未知参数 $c_1, c_2, \cdots, c_8$。这些参数就组成了一个模型，该模型把一幅图像的像素位置变换为另一幅需要配准的图像的像素位置。

111 ~ 112

一旦有了系数，式(2.6-24)和式(2.6-25)就成为我们变换输入图像中的所有像素并生成期望新图像的工具，如果约束点选择得正确，新图像就能与参考图像配准。在 4 个约束点不足以得到令人满意的

配准的情况下，常常使用的一种方法是选择大量的约束点，如同一个子图像那样，用一组 4 个约束点形成四边形来处理。子图像用上述方法处理，四边形内部的所有像素使用由这些约束点求出的系数进行变换。然后，移动到另一组 4 约束点，重复上述过程，直到所有四边形区域均被处理为止。当然，也可使用比四边形更复杂的区域并采用更复杂的模型，例如使用最小均方算法的拟合多项式。通常，控制点的数量和解决问题所要求的模型的复杂程度取决于几何畸变的严重程度。最后，应牢记式(2.6-24)和式(2.6-25)所定义的变换，或任何针对该问题的其他模型，以及输入图像中像素空间坐标的简单映射。我们还需要使用前面讨论的方法执行灰度内插，以便给这些像素灰度值赋。

例 2.10　图像配准。

图2.37(a)显示了一幅参考图像，图2.37(b)显示了相同的图像，但因垂直和水平切变产生了几何畸变。我们的目标是使用参考图像得到约束点，然后用约束点配准这两幅图像。我们(手工)选择的约束点靠近图像的角点，由白色小方形点表示(因为畸变是在两个方向上的线性拉伸，因此仅需要 4 个约束点)。图 2.37(c)显示了使用前面讨论过的这些约束点来实现配准的结果。我们注意到，配准并不完美，因为图 2.37(c)中有明显的黑边。图2.37(d)更清楚地显示了参考图像和校正后的图像间配准的细小不足。导致这种差别的原因是手工选择约束点时存在误差。畸变很严重时，使用约束点很难实现完美的匹配。

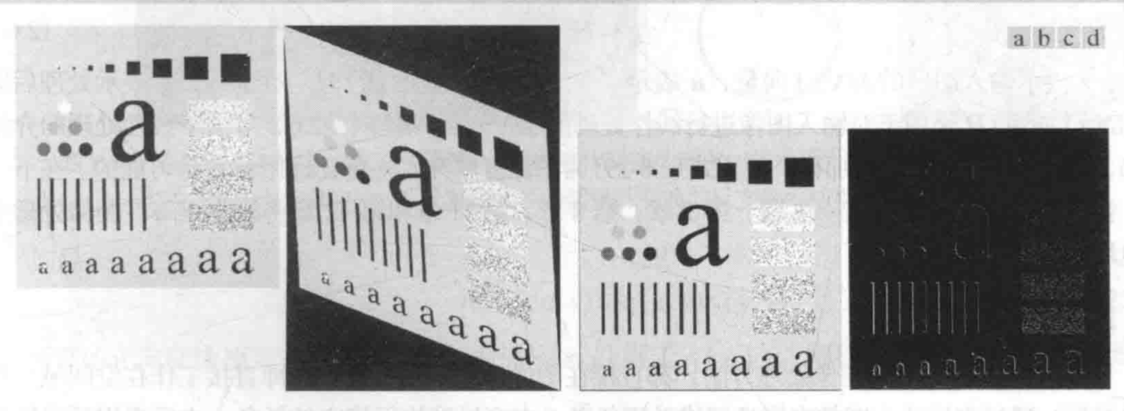


图 2.37　图像配准：(a)参考图像；(b)输入图像(几何畸变图像)，相应的约束点以靠近角点的白色小方块显示；(c)配准后的图像(注意边界上的误差)；(d)图(a)和图(c)的差，显示了更多的配准误差

2.6.6　向量与矩阵操作

多光谱图像处理是使用向量和矩阵操作的典型领域。例如，在第 6 章中，我们将介绍在 RGB 彩色空间中如何使用红色、绿色、蓝色分量图像来形成如图 2.38 说明的彩色图像。这里，我们看到 RGB 图像的每个像素都有三个分量，这些分量可组织成一个列向量的形式：

> 关于向量和矩阵的简单介绍，请参阅本书网站上的 Tutorial 部分。

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \tag{2.6-26}$$

式中，z_1 是红色图像中像素的亮度，其他两个元素是相应的绿色图像和蓝色图像中像素的亮度。这样，大小为 $M \times N$ 的 RGB 彩色图像就可用这一大小的三个分量图像来表示，或总共用 MN 个三维向量来表示。普通的多光谱情况包括 n 幅分量图像(见图 1.10)，将形成 n 维向量。在第 6 章、第 10 章、第 11 章和第 12 章，我们都使用这种形式的向量表示。

一旦像素表示为向量，我们就可以使用向量矩阵理论这一工具。例如，一个像素向量 $\mathbf{z}$ 和一个任意点 $\mathbf{a}$ 在 n 维空间的欧氏距离 D 可以用一个向量积来定义：

112 ~ 114

$$D(\boldsymbol{z},\boldsymbol{a})=\left[(\boldsymbol{z}-\boldsymbol{a})^{\mathrm{T}}(\boldsymbol{z}-\boldsymbol{a})\right]^{\frac{1}{2}}=\left[(z_1-a_1)^2+(z_2-a_2)^2+\cdots+(z_n-a_n)^2\right]^{\frac{1}{2}} \tag{2.6-27}$$

我们看到，这是式(2.5-1)定义的二维欧氏距离的通式。式(2.6-27)有时称为向量范数，表示为$\|\boldsymbol{z}-\boldsymbol{a}\|$。在后面的章节中，我们将多次使用距离计算。

像素向量的另一个重要优点是由下式表示的线性变换：

$$\boldsymbol{w}=\boldsymbol{A}(\boldsymbol{z}-\boldsymbol{a}) \tag{2.6-28}$$

式中，$\boldsymbol{A}$是大小为$m\times n$的矩阵，$\boldsymbol{z}$和$\boldsymbol{a}$是$n\times 1$大小的列向量。如后面所述，这种类型的变换在图像处理中的用途很多。

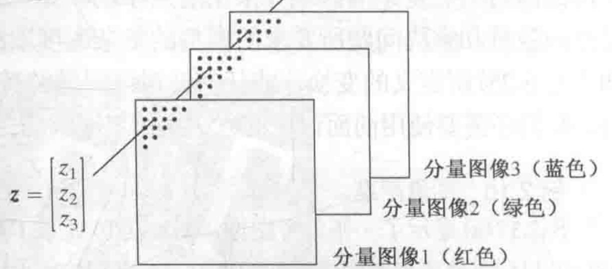


图 2.38 三幅 RGB 分量图像中的相应像素值形成一个向量

如式(2.4-2)表示的那样，整个图像可当作矩阵(或向量)来处理，这在求解大量数字图像处理问题时意义重大。例如，我们可以把一幅大小为$M\times N$的图像描述为一个$MN\times 1$维向量，方法是用图像的第一行作为该向量的前N个元素，第二行作为下N个元素，等等。用这种方法形成的图像，我们可以表达应用更广范围如下图像线性处理：

$$\boldsymbol{g}=\boldsymbol{Hf}+\boldsymbol{n} \tag{2.6-29}$$

式中，$\boldsymbol{f}$表示输入图像的$MN\times 1$向量，$\boldsymbol{n}$表示一个$M\times N$噪声模式的$MN\times 1$向量，$\boldsymbol{g}$表示处理后图像的$MN\times 1$向量，$\boldsymbol{H}$是用于对输入图像进行线性处理的$MN\times MN$矩阵(见2.6.2节关于线性处理的介绍)。例如，像5.9节将要讨论的那样，使用式(2.6-29)为图像复原开发一整套通用技术是可能的。在下一节中，我们会再次接触矩阵这一主题，第5章、第8章、第11章和第12章将给出矩阵在图像处理中的其他用途。

2.6.7 图像变换

目前为止讨论的所有图像处理方法，都直接在图像像素上进行操作，即直接工作在空间域。在有些情况下，通过变换输入图像来表达图像处理任务，在变换域执行指定的任务，之后再用反变换返回到空间域，更为可取。随着学习的深入，我们会遇到许多不同的变换。表示为$T(u,v)$的二维线性变换是一类特别重要的变换，其通用形式可表达为

$$T(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)r(x,y,u,v) \tag{2.6-30}$$

式中，$f(x,y)$是输入图像，$r(x,y,u,v)$称为正变换核，式(2.6-30)对$u=0,1,2,\cdots,M-1$和$v=0,1,2,\cdots,N-1$进行计算。与以前一样，x和y是空间变量，M和N是f的行和列，u和v称为变换变量。$T(u,v)$称为$f(x,y)$的正变换。给定$T(u,v)$后，我们可以用$T(u,v)$的反变换还原$f(x,y)$：

$$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}T(u,v)s(x,y,u,v) \tag{2.6-31}$$

式中，$x=0,1,2,\cdots,M-1$，$y=0,1,2,\cdots,N-1$，$s(x,y,u,v)$称为反变换核。式(2.6-30)和式(2.6-31)一起称为变换对。

图2.39显示了在线性变换域执行图像处理的基本步骤。首先，变换输入图像，然后用预定义的操作修改该变换，最后，输出图像由计算修改后的变换的反变换得到。这样，我们可以看出，该过程先从空间域到变换域，然后返回空间域。

114 ~ 116

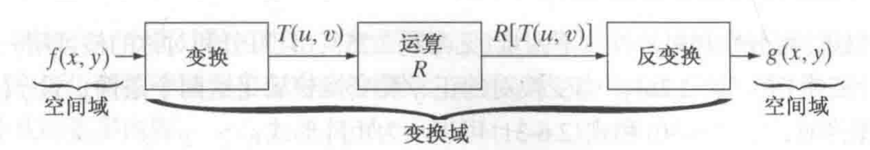


图 2.39 线性变换域中操作的一般方法

例 2.11 变换域图像处理。

图2.40 显示了图2.39 中步骤的一个例子。在这种情况下，所用的变换是傅里叶变换，在本节稍后我们将简单提一下它，第 4 章将对其进行详细介绍。图2.40(a)是一幅被正弦干扰污染了的图像，图2.40(b)是该图像的傅里叶变换的幅度，它是图2.39 中第一阶段的输出。如第 4 章所示，空间域中的正弦干扰在变换域以较亮的脉冲方式出现。在这种情况下，脉冲是图 2.40(b)中可见的圆形模式。图 2.40(c)显示了一个模板图像(称为滤波器)，白和黑分别代表 1 和 0。对于该例，图 2.39 中第二个方框内的操作是变换的结果与模板相乘，以可靠地消除干扰脉冲。图2.40(d)显示了最终结果，该结果是通过计算修改后的变换的反变换得到的。事实上，我们恰好可以看到基准标志(淡淡的十字)，它用于图像对准。

a b c d

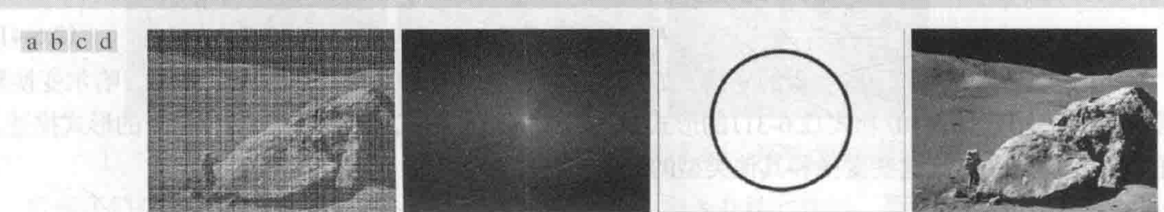


图 2.40 (a)被正弦干扰污染了的图像；(b)傅里叶变换的幅度，显示了干扰能量脉冲；(c)用于消除能量脉冲的模板；(d)计算修改后的傅里叶变换的反变换的结果(原图像由 NASA 提供)

如果

$$r(x, y, u, v) = r_1(x, u) r_2(y, v) \tag{2.6-32}$$

那么所谓的正向变换核是可分的。另外，如果 $r_1(x, y)$ 等于 $r_2(x, y)$，则称变换核是对称的，从而有

$$r(x, y, u, v) = r_1(x, u) r_1(y, v) \tag{2.6-33}$$

在前面的公式中，若用 s 代替 r，则同样的说明适用于反变换核。

例 211 中讨论的二维傅里叶变换有如下正、反变换核：

$$r(x, y, u, v) = \mathrm{e}^{-\mathrm{j}2\pi(ux/M + vy/N)} \tag{2.6-34}$$

和

$$s(x, y, u, v) = \frac{1}{MN}\mathrm{e}^{\mathrm{j}2\pi(ux/M + vy/N)} \tag{2.6-35}$$

式中，$\mathrm{j} = \sqrt{-1}$，因此这些核是复数。将这些核代入式(2.6-30)和式(2.6-31)给出的通用变换公式中，可得出离散傅里叶变换对：

$$T(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x, y)\mathrm{e}^{-\mathrm{j}2\pi(ux/M + vy/N)} \tag{2.6-36}$$

和

$$f(x, y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} T(u,v)\mathrm{e}^{\mathrm{j}2\pi(ux/M + vy/N)} \tag{2.6-37}$$

从基础意义上说，这些公式在数字图像处理中很重要，在第 4 章中，我们将投入更多的篇幅从基本原理入手推导它们，然后在较宽的应用范围使用它们。

116 ~ 117

证明傅里叶核的可分性和对称性并不困难(见习题2.25)，且可分和对称的核可用一维傅里叶变换计算二维傅里叶变换(见习题2.26)。当变换对的正、反变换核满足这两个条件，且 $f(x,y)$ 是大小为 $M \times M$ 的方形图像时，式(2.6-30)和式(2.6-31)可表示为矩阵形式：

$$\boldsymbol{T} = \boldsymbol{AFA} \tag{2.6-38}$$

式中，$\boldsymbol{F}$ 是包含元素 $f(x,y)$ 的 $M \times M$ 矩阵［见式(2.4-2)］；$\boldsymbol{A}$ 是有元素 $a_{ij} = r_1(i,j)$ 的 $M \times M$ 矩阵，$\boldsymbol{T}$ 是 $M \times M$ 变换的结果，其值为 $T(u,v)$, $u, v = 0, 1, 2, \cdots, M-1$。

为得到反变换，我们用反变换矩阵 $\boldsymbol{B}$ 前乘和后乘式(2.6-38)：

$$\boldsymbol{BTB} = \boldsymbol{BAFAB} \tag{2.6-39}$$

如果 $\boldsymbol{B} = \boldsymbol{A}^{-1}$，则

$$\boldsymbol{F} = \boldsymbol{BTB} \tag{2.6-40}$$

该式指出 $\boldsymbol{F}$［其所有元素均等于 $f(x,y)$］可完全由其正向变换复原。如果 $\boldsymbol{B} \neq \boldsymbol{A}^{-1}$，则用式(2.6-40)可得到一个近似：

$$\hat{\boldsymbol{F}} = \boldsymbol{BAFAB} \tag{2.6-41}$$

除傅里叶变换之外，一些重要的变换，如沃尔什变换、哈达玛变换、离散余弦变换、哈尔变换和斜变换等，可用式(2.6-30)和式(2.6-31)的形式描述，或等价地用式(2.6-38)和式(2.6-40)的形式描述。我们将在后续章节讨论这些变换和其他类型的图像变换。

2.6.8 概率方法

> 关于概率论的简介，请读者查阅本书网站上的 Tutorials 部分。

概率以很多方式用于图像处理工作中。最简单的方式是当我们以随机量处理灰度值时。例如，令 $z_i, i = 0, 1, 2, \cdots, L-1$ 表示一幅 $M \times N$ 大小数字图像中所有可能的灰度值，则在给定图像中灰度级 z_k 出现的概率 $p(z_k)$ 可估计为

$$p(z_k) = \frac{n_k}{MN} \tag{2.6-42}$$

式中，n_k 是灰度 z_k 在图像中出现的次数，MN 是像素总数。显然，

$$\sum_{k=0}^{L-1} p(z_k) = 1 \tag{2.6-43}$$

一旦我们知道了 $p(z_k)$，就可以得出许多重要的图像特性。例如，平均灰度由下式给出：

$$m = \sum_{k=0}^{L-1} z_k p(z_k) \tag{2.6-44}$$

类似地，灰度的方差是

$$\sigma^2 = \sum_{k=0}^{L-1} (z_k - m)^2 p(z_k) \tag{2.6-45}$$

方差是 z 值关于均值的展开度的度量，因此它是图像对比度的有用度量。通常，随机变量 z 关于均值的第 n 阶矩定义为

$$\mu_n(z) = \sum_{k=0}^{L-1} (z_k - m)^n p(z_k) \tag{2.6-46}$$

118 ~ 119

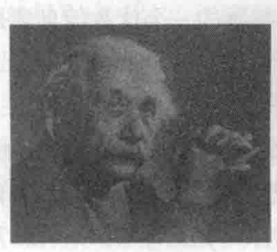
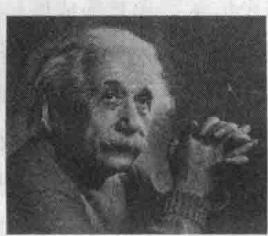
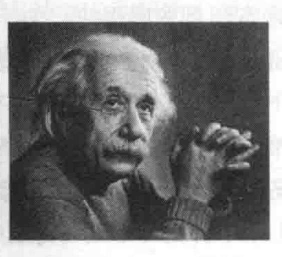
我们看到，$\mu_0(z) = 1, \mu_1(z) = 0$ 且 $\mu_2(z) = \sigma^2$。反之，均值和方差与图像的视觉特性有明显的直接关系，

高阶矩更敏感。例如，一个正三阶矩指出其灰度倾向于比均值高，负三阶矩则倾向于比均值低，而零三阶矩告诉我们灰度近似相等地分布在均值的两侧。这些特性对于计算目的很有用，但它们一般不能告诉我们图像外观的更多内容。

例 2.12　图像灰度对比度度量时标准差值的比较。

图2.41 显示了分别展示低、中、高对比度的三幅 8 比特图像。在这三幅图像中，像素灰度的标准差分别是 14.3, 31.6 和 49.2 个灰度级。相应的方差值分别是 204.3, 997.8, 2424.9。两组数值告诉我们相同的内容，但这三幅图像可能的灰度值范围是[0, 255]，与该范围相关的标准差值与方差相比要直观得多。

> 方差的单位是灰度值的平方。在比较对比度值时，我们通常替代使用标准差 σ（方差的均方根），因为其维数直接取决于灰度值。

a b c

图 2.41　展示了(a)低对比度、(b)中对比度和(c)高对比度的图像

后面我们会看到，在开发图像处理算法时，概率这一概念起核心作用。例如，在第 3 章，我们利用式(2.6-42)的概率度量推导灰度变换算法。在第 5 章中，我们使用概率和矩阵公式开发图像复原算法。在第 10 章中，我们使用概率来分割图像，而在第 11 章中，我们用概率来描述纹理。在第 12 章中，我们以概率公式为基础，导出最佳目标识别技术。

到目前为止，我们已经讨论了将概率应用到单幅二维图像上的单个随机变量(灰度)的问题。考虑序列图像，我们可将时间作为第三个变量。处理这一复杂性的工具是随机图像处理技术。我们可以进一步将整个图像考虑为一个空间随机事件(相对于一个点)。基于这一概念的处理工具来自于随机域技术。在 5.8 节中，我们将给出一个例子，说明如何像随机事件那样处理整个图像，但关于随机过程和随机域的讨论超出了本书的范围。本节末尾提供的参考文献是阅读这些主题的起点。

小结

本章内容为后续讨论提供了主要的背景知识。对人类视觉系统的论述虽然简单，但提供了人眼感知图像信息能力的基本概念。关于光和电磁波谱的讨论是理解本书所用许多图像来源的基础。类似地，2.3.4 节给出的图像模型也是第 4 章中称为同态滤波图像增强技术的基础。

2.4 节介绍的取样和内插的概念，是读者在实践中可能遇到的许多数字现象的基础。掌握傅里叶变换和频率域的概念后，在第 4 章中我们还将回到取样及其分支问题的讨论。

2.5 节介绍的概念是基于像素邻域处理技术的基本组成部分。例如，如第 3 章和第 5 章所示的那样，邻域处理方法是许多图像增强和复原过程的核心。在第 9 章中，我们将采用邻域操作对图像做形态学处理；在第 10 章中，我们将用邻域处理分割图像；在第 11 章中，我们用它们描述图像。在应用方面，邻域处理由于处理速度快，且硬件和/或固件实现简单，故在图像处理商业应用方面得到了人们的偏爱。

2.6 节的内容对于全书的阅读很有帮助，虽然讨论是介绍性的，但可帮助读者建立数字图像处理

119 ~ 120

的概念。如该节中提到的那样，介绍的那些工具在后续章节中都会用到。单独一章或附录无法充分论述数学概念，因此后续章节会深入介绍 2.6 节的数学工具，并说明如何运用它们来解决图像处理问题。

参考文献

2.1 节中有关人眼结构的内容见 Atchison and Smith[2000]和 Oyster[1999]。阅读资料见 Regan[2000]和 Gordon[1997]。Hubel[1988]和 Cornsweet[1970]也很重要。Born and Wolf[1999]是讨论光和电磁波理论的基本参考资料。Felsen and Marcuvitz[1994]详细论述了电磁波的传播。

图像感知领域非常宽泛且发展很快。关于光传感器和其他成像传感器的一个优秀信息源是国际光学工程学会(SPIE)。下面是 SPIE 在该领域的代表性出版物：Blouke et al.[2001]、Hoover and Doty[1996]和 Freeman[1987]。

2.3.4 节给出的图像模型来自 Oppenheim, Schafer and Stockham[1968]。该节中所用的照度值和反射值的参考文献是 *IESNA Lighting Handbook*[2000]。关于图像取样及其一些效应(譬如混淆效应)的其他读物见 Bracewell[1995]。第 4 章中将更详细地讨论这一主题。2.4. 3 节中作为取样和量化的函数的感知图像质量方面的早期实验见 Huang[1965]。在降低图像取样数和灰度级的同时，尽量减少随之而来的退化仍是当前人们感兴趣的主题，详见Papamarkos and Atsalakis[2000]。关于图像收缩和放大方面的深入读物见Sid-Ahmed[1995], Unser et al.[1995], Umbaugh[2005]和 Lehmann et al.[1999]。2.5 节涵盖主题的深入读物见 Rosenfeld and Kak[1982], Marchand-Maillet and Sharaiha [2000]和 Ritter and Wilson[2001]。

图像处理中关于线性系统(见 2.6.2 节)的额外读物见 Castleman[1996]。使用图像平均来消除噪声的方法(见 2.6.3 节)由 Kohler and Howell[1963]首次提出。关于随机变量求和的均值和方差的期望值的材料见 Peebles[1993]。图像相减(见 2.6.3 节)是广泛用于检测变化的通用图像处理工具。为使得图像相减有意义，必须配准被减图像，或必须识别出由于移动导致的人为缺陷。由 Meijering et al.[1999, 2001]撰写的两篇论文给出了实现这些目的的技术。

2.6.4 节的基本参考文献是 Cameron[2005]。这一主题的高级读物见Tourlakis[2003]。关于模糊集合的介绍见 3.8 节和第 3 章中的相关参考文献。单点和邻域处理(见 2.6.5 节)的细节见 3.2 节到 3.4 节及第 3 章中关于这些主题的参考文献。关于几何空间变换的内容，见 Wolberg[1990]。

Noble and Daniel[1988]是矩阵和向量操作(2.6.6 节)的基本参考材料。傅里叶变换(见 2.6.7 节)的详细讨论见第 4 章，数字图像处理中使用的其他类型变换例子见第 7 章、第 8 章和第 11 章。概率和随机变量(见 2.6.8 节)的基本介绍见 Peebles[1993]，深入介绍见 Papoulis[1991]。随机和随机域图像处理的基础材料，见 Rosenfeld and Kak[1982], Jähne[2002]和 Won and Gray[2004]。

本章中给出的许多技术的软件实现细节见 Gonzalez, Woods and Eddins[2004]。

习题

标有星号的习题的详细答案可在本书的网站上找到。该网站上还包含有基于本章内容建议的项目。

★2.1 使用 2.1 节提供的背景信息，如果纸上的一个打印点离眼睛 0.3 m 远，请采用纯几何方法，估计眼睛能辨别的最小打印点的直径。为简单起见，假设在中央凹处的像点变得远比视网膜区域的感受器(锥状体)的直径小时，视觉系统已不能检测到该点。进一步假设中央凹可建模为直径为1.5 mm 的圆形阵列，并且锥状体间的间隔在该阵列上均匀分布。

120 ~ 121

2.2 当我们白天进入一家黑暗的剧场时，在能看清并找到空座时要用一段时间适应。2.1 节描述的视觉过程在这种情况下起什么作用？

★**2.3** 虽然图 2.10 中未显示，但交流电的确是电磁波谱的一部分。美国的商用交流电频率是 77 Hz。问这一波谱成分的波长是多少米？

2.4 设计一个研究细胞、细菌、病毒和蛋白质边界形状的图像系统前端。在这种情况下，前端由光源和相应的图像摄像机组成。要求围成各个类别的圆形的直径分别为 25 μm, 0.5 μm, 0.05 μm 和 0.005 μm。

(a) 能用单个传感器和摄像机解决这一成像问题吗？如果回答是肯定的，试给出照射波长和所需的摄影机类型。这里的“类型”指的是摄像机最敏感的电磁波谱波段(譬如红外波段)。

(b) 如果(a)中你的回答是否定的，那么你建议采用哪种类型的照明光源和相应的图像传感器？指出(a)中所要求的光源和摄像机。要用最少数量的照明光源和摄像机解决这个问题。“解决这个问题”是指能够分别检测直径为 25 μm, 0.5 μm, 0.05 μm 和 0.005 μm 的环形细节。

2.5 14 mm×14 mm 的 CCD 摄像机芯片有 2048×2048 个元素，将它聚焦到相距 0.5 m 远的一个方形平坦区域。该摄像机每毫米能分辨多少线对？摄像机配备了一个 35 mm 镜头。(提示：成像处理模型如图 2.3 所示，但使用摄像机镜头的焦距替代眼睛的焦距。)

★**2.6** 某汽车制造商想在一条限量版跑车生产线上将一些零件自动安放到跑车的保险杠上。这一部件就是彩色坐标，为选择合适的保险杠零件，机器人需要知道每辆车的颜色。模型有 4 种颜色：蓝色、绿色、红色和白色。请提出一种基于成像的解决方案。你怎样解决自动确定每辆跑车颜色的问题？记住，在选择零件时价格是最主要的考虑因素。

2.7 假设中心在 (x_0, y_0) 的平坦区域被一个强度分布为

$$i(x,y) = K\mathrm{e}^{-\left[(x-x_0)^2+(y-y_0)^2\right]}$$

的光源照射。为简单起见，假设该区域的反射是恒定的并且等于 1.0，令 $K=255$。如果图像使用 k 比特的强度分辨率进行数字化，且眼睛可检测相邻像素间的 4 种灰度突变，请问 k 取何值时会出现可见的伪轮廓？

2.8 画出习题 2.7 中 $k=4$ 时的图像。

★**2.9** 数字数据传输通常用波特率度量，其定义为每秒钟传输的比特数。通常，传输是以一个开始比特、一个字节(8 比特)的信息和一个停止比特组成的包完成的。利用这些事实，回答下列问题：

(a) 使用 33.6K 波特的调制解调器传输一幅大小为 2048×2048 的 256 灰度级的图像，需要几分钟？

(b) 波特率为 3000K 时，即典型电话 DSL(数字用户线)的媒介速率时，传输要用多长时间？

2.10 高清晰度电视(HDTV)使用 1080 条水平电视线隔行扫描来产生图像(每隔一行在显像管表面画一条线，每两场形成一帧，每场用时 1/60 秒)。图像的宽高比是 16 : 9。在水平行数不变的情况下，求图像的垂直分辨率。一家公司已经设计了一种图像获取系统，该系统由 HDTV 图像生成数字图像。在该系统中，每条(水平)电视行的分辨率与图像的宽高比成正比，彩色图像的每个像素都有 24 比特的灰度分辨率，红色、绿色、蓝色图像各 8 比特。这三幅原色图像形成彩色图像。存储 90 分钟的一部 HDTV 电影需要多少比特？

★**2.11** 考虑两个图像子集 S_1 和 S_2，如下页右图所示。对于 $V=\{1\}$，确定这两个子集是(a)4 邻接的、(b)8 邻接的，还是(c)m 邻接的？

★**2.12** 提出将像素宽度的 8 通路转换为 4 通路的一种算法。

122 ~ 123

2.13 提出将像素宽度的 m 通路转换为 4 通路的一种算法。

2.14 参阅 2.5.2 节末尾的讨论，在该讨论中，我们将背景定义为 $(R_u)^c$，即图像中所有区域的并集的补集。在某些应用中，定义一个像素 $(R_u)^c$ 的子集为背景是有利的，该子集不是孔洞像素区域(非正式地，我们认为孔洞是由区域像素围绕的背景像素子集)。如何修改这一定义，以便从 $(R_u)^c$ 中排除孔洞像素？像“背景是非孔洞像素的 $(R_u)^c$ 像素的子集”这样的答案是不能接受的。(提示：利用连通性这一概念。)

	S_1				S_2				
0	0	0	0	0	0	0	1	1	0
1	0	0	1	0	0	1	0	0	1
1	0	0	1	0	1	1	0	0	0
0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	0	0	1	1	1

2.15 考虑右图的图像分割。

★(a) 令 $V=\{0,1,2\}$，计算 p 和 q 间 4、8 和 m 通路的最短长度。如果在这两点间不存在一个特殊通路，试解释原因。

(b) 令 $V=\{2,3,4\}$，重复问题(a)。

3	4	1	2	0
0	1	0	4	2 (q)
2	2	3	1	4
(p) 3	0	4	2	1
1	2	0	3	4

2.16 ★(a) 对于点 p 和 q 间的 D_4 距离等于这两点间最短 4 通路的情况，给出需要的条件。

(b) 这个通路唯一吗？

2.17 对于 D_8 距离，重做习题 2.16。

★2.18 下一章中将讨论一些算子，它们的功能是在一个很小的子图像区域 S 中计算像素值的总和。证明这些算子都是线性算子。

2.19 一个数集的中值 ξ 定义为这样一个值，数集中的一半数值比它小，另一半数值比它大。例如，数集 $\{2, 3, 8, 20, 21, 25, 31\}$ 的中值是 20。试证明计算子图像区域 S 的中值的算子是非线性的。

★2.20 证明式(2.6-6)和式(2.6-7)的正确性。[提示：从式(2.6-4)开始，并利用这样一个事实：一个和的期望值就是期望值的和。]

2.21 考虑两幅 8 比特图像，它们的灰度级跨越从 0 到 255 的全部范围。

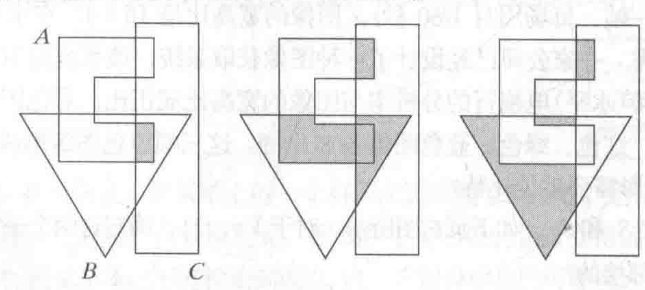
(a) 讨论从图像(1)中重复减去图像(2)的限制效应。假设结果仍然用 8 比特表示。

(b) 如果颠倒图像的顺序，会产生不同的结果吗？

★2.22 图像相减通常用于在生产线上检测缺失的元件。方法是存储一幅对应于正确组装的“金”图像；然后，从相同产品的传入图像中减去该图像。理想情况下，如果新产品组装正确，则差值应为零。对于缺失元件的产品，在不同于金图像的位置，差值图像将不为零。在实际中，你认为在什么条件下使用这种方法工作才是合适的？

2.23 ★(a) 参考图 2.31，画出集合 $(A^c - B)\cup(B - A)$。

(b) 给出下图中集合 A, B, C 中所示的阴影集合的表达式。每幅图中的阴影区域组成一个集合，因此应该对这三幅图的每个阴影区域给出一个表达式。



123 ~ 124

2.24 使用三角形区域替代四边形区域时，所得到的类似于式(2.6-24)和式(2.6-25)的等式是什么？

2.25 证明式(2.6-34)和式(2.6-35)中的傅里叶核是可分的和对称的。

★**2.26**　证明具有可分性、对称性的二维变换核可以用(1)沿输入的行(列)的一维变换来计算；接着(2)用第一步结果的列(行)的一维变换来计算。

2.27　一家工厂生产半透明的微型聚合物方形产品。要求进行 100%的视觉检测来严格控制质量，工厂管理者发现人工检测越来越贵，于是检测改为半自动方式。在每个检测位置，机械装置把聚合物方形产品放在一盏灯的上方，灯位于光学系统的下方，该光学系统产生方形产品的放大图像，图像完全落在大小为 80 mm × 80 mm 的观察屏上。缺陷以暗斑形式出现，检测者的工作是观察屏幕。若屏幕上一个样品中出现了多个直径大于等于 0.8 mm 的黑斑，则丢弃该样品。管理者相信，如果能找到一种全自动的处理方法，利润将提高 50%。她还相信项目的成功是她升迁为法人的阶梯。经过大量调研后，管理者确定了解决这一问题的方法，即用 CCD 电视摄像机去观察每个检测屏幕，并把摄像机的输出送到能够检测斑点的图像处理系统中，测量它们的直径，并触发以前由检测员操作的接受和丢弃按钮。她可以找到一个系统，只要最小的缺陷在数字图像中至少为 2×2 像素的区域发生，就可以完成该工作。假设管理者聘任你来帮助她确定摄像机和镜头系统的参数，但要求你使用现有的元件。对于镜头，假设该约束条件意味着任何 25 mm、35 mm 直至 200 mm 焦距的整数倍。对于摄像机，这意味着分辨率为 512 × 512 像素、1024 × 1024 像素或 2048 × 2048 像素。这些摄像机中的各个成像元素是大小为 8 μm × 8 μm 的方形，成像元素的间距为 2 μm。对于该应用，摄像机要远贵于镜头，因此该问题应以选择镜头为基础，尽可能使用最低分辨率的摄像机来解决。作为一名顾问，请你提供一份书面建议，对你的结论给出合理且详细的分析。使用习题 2.5 中建议的相同成像几何。

124～125



第3章　灰度变换与空间滤波

It makes all the difference whether one sees darkness through the light or brightness through the shadows.

David Lindsay

引言

术语空间域指图像平面本身，这类图像处理方法直接操作图像中的像素。这是相对于变换域中的图像处理而言的，如2.6.7节所介绍的和第4章将要详细讨论的那样，变换域的图像处理首先把一幅图像变换到变换域，在变换域中进行处理，然后通过反变换把处理结果返回到空间域。空间域处理主要分为灰度变换和空间滤波两类。如本章介绍的那样，灰度变换对图像的单个像素进行操作，主要以对比度和阈值处理为目的。空间滤波涉及改善性能的操作，如通过图像中每个像素的邻域处理来锐化图像。接下来几节讨论一些经典的灰度变换和空间滤波技术。还将讨论模糊技术的某些细节，以便在灰度变换和空间滤波算法的公式化表示中，并入不太精确的信息。

3.1　背景知识

3.1.1　灰度变换和空间滤波基础

本节讨论的所有图像处理技术都是在空间域进行的。由2.4.2节的讨论，我们知道空间域是包含图像像素的简单平面。如2.6.7节定义的那样，与频率域（第4章的主题）相反，空间域技术直接操作图像的像素。例如，频率域的操作在图像的傅里叶变换上执行，而不是在图像本身上执行。如我们阅读全书过程中将要了解到的那样，某些图像处理任务在空间域中执行更容易或更有意义，而另一些任务则更适合使用其他方法。通常，空间域技术在计算上更有效，且执行所需的处理资源较少。

本章讨论的空间域处理可由下式表示：

$$g(x,y)=T[f(x,y)] \tag{3.1-1}$$

式中，$f(x,y)$是输入图像，$g(x,y)$是处理后的图像，T是在点(x,y)的邻域上定义的关于f的一种算子。算子可应用于单幅图像（本章中我们的主要关注点）或图像集合，例如，如2.6.3节中讨论的那样，为降低

有时也使用其他形状的邻域，如圆的数字近似，但矩形邻域是到目前为止最好的邻域，因为它们在计算机上实现起来更为容易。

噪声而对图像序列执行逐像素的求和操作。图3.1显示了式(3.1-1)在单幅图像上的基本实现。所示的点 (x, y) 是图像中的一个任意位置，包含该点的小区域是点 (x, y) 的邻域，如2.6.5节解释的那样。邻域通常是中心在 (x, y) 的矩形，其尺寸要比图像小得多。

图 3.1 中给出的处理由以下几步组成：邻域原点从一个像素向另一个像素移动，对邻域中的像素应用算子 T，并在该位置产生输出。这样，对于任意指定的位置 (x, y)，输出图像 g 在这些坐标处的值，等于对 f 中以 (x, y) 为原点的邻域应用算子 T 的结果。例如，假设该邻域是大小为 3×3 的正方形，算子 T 定义为“计算该邻域的平均灰度”。考虑图像中的任意位置，譬如(100,150)。假设该邻域的原点位于其中心处，则在该位置的结果 $g(100,150)$ 是计算 $f(100,150)$ 和其 8 个邻点的和，再除以 9（即由邻域包围的像素灰度的平均值）。然后，邻域的原点移到下一个位置，并重复前面的过程，产生下一个输出图像 g 的值。该处理通常从输入图像的左上角开始，以水平扫描的方式逐像素地处理，每次一行。当该邻域的原点位于图像的边界上时，部分邻域将位于图像的外部。此时，不是在用 T 做指定的计算时忽略外侧邻点，就是用 0 或其他指定的灰度值填充图像的边缘。被填充边界的厚度取决于邻域的大小。在 3.4.1 节中，我们将再回到这一问题。

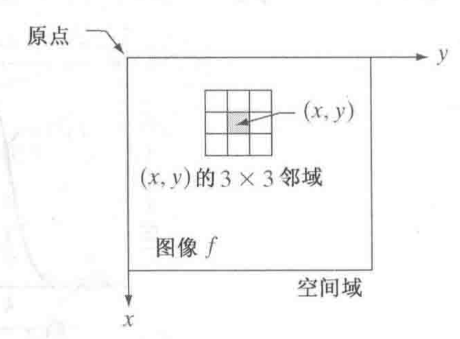


图 3.1　空间域一幅图像中关于点 (x, y) 的一个 3×3 邻域。邻域在图像中从一个像个素到另一个像素移动来生成一幅输出图像

就像在 3.4 节中详细讨论的那样，刚刚描述的过程称为空间滤波，其中邻域与预定义的操作一起称为空间滤波器(也称为空间掩模、核、模板或窗口)。在邻域中执行的操作类型决定了滤波处理的特性。

最小邻域的大小为 1×1。在这种情况下，g 仅取决于点 (x, y) 处的 f 值，而式(3.1-1)中的 T 则成为一个形如下式的灰度(也称为灰度级或映射)变换函数：

$$s = T(r) \tag{3.1-2}$$

式中，为表达方便，令 r 和 s 分别表示变量，即 g 和 f 在任意点 (x, y) 处的灰度。例如，如果 $T(r)$ 有如图3.2(a)所示的形式，对 f 中每个像素施以变换产生相应 g 的像素的效果，将比原始图像有更高的对比度，即低于 k 的灰度级更暗，而高于 k 的灰度级更亮。这种技术有时称为对比度拉伸(见3.2.4节)，在该技术中，低于 k 的 r 值被变换函数压缩在一个较窄的范围 s 内，接近黑色；而高于 k 的 r 值则与此相反。很明显，灰度值 r_0 经映射得到了相应的值 s_0。在如图3.2(b)所示的极限情况下，$T(r)$ 产生了一幅两级(二值)图像。这种形式的映射称为阈值处理函数。有些相当简单但功能强大的处理方法，可以使用灰度变换函数用公式加以表达。本章主要使用灰度变换函数来增强图像。第 10 章将使用灰度变换函数来分割图像。结果仅取决于一个点的灰度的方法，有时称为点处理技术，与此相反，本节早些时候讨论的方法则称为邻域处理技术。

3.1.2　关于本章中的例子

虽然灰度变换和空间滤波覆盖了相当宽的应用范围，但本章中的多数例子是图像增强应用。增强处理是对图像进行加工，使其结果对于特定的应用比原始图像更合适的一种处理。“特定”一词在这里很重要，它一开始就确定增强技术是面向问题的。例如，对于增强 X 射线图像非常有用的方法，可能并不是增强由电磁波谱中远红外波段拍摄的图像的最好方法。图像增强没有通用的“理论”。当为视觉解释而处理一幅图像时，观察者将是判定一种特定方法好坏的最终裁判员。在处理机器感知

127 ~ 128

时，一种给定的技术很易于量化。例如，在自动字符识别系统中，最合适的增强方法是能得到最好识别率的方法，这里不考虑一种方法较另一种方法的计算量的要求。

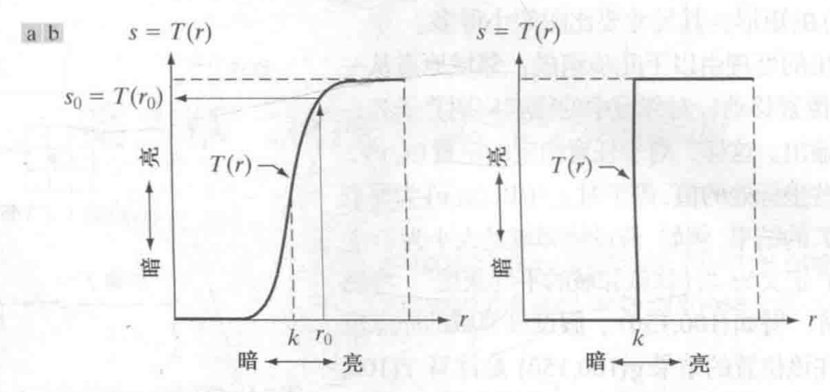


图 3.2 灰度变换函数：(a)对比度拉伸函数；(b)阈值处理函数

然而，不管应用或使用的方法是什么，图像增强都是视觉上最具吸引力的图像处理领域之一。理所当然地，图像处理的初学者通常会寻找重要且理解起来相对简单的增强应用。因此，使用图像增强的例子来说明本章中论述的空间处理方法，不仅可以节省本书中处理图像增强的额外章节，而且更重要的是，是对初学者介绍空间域处理技术细节的更有效方法。如我们将要看到的那样，随着本书的阅读进度，本章叙述的基本内容可用于比图像增强宽得多的范围。

3.2 一些基本的灰度变换函数

灰度变换是所有图像处理技术中最简单的技术。r 和 s 分别代表处理前后的像素值。如前一节指出的那样，这些值与 $s=T(r)$表达式的形式有关，其中T 是把像素值r 映射到像素值s 的一种变换。由于我们处理的是数字量，所以变换函数的值通常存储在一个一维阵列中，且从r 到s 的映射是通过查找表实现的。对于 8 比特环境，包含T 的值的一个查找表将有 256 条记录。

作为关于灰度变换的介绍，考虑图 3.3，该图显示了图像增强常用的三类基本函数：线性函数(反转和恒等变换)、对数函数(对数和反对数变换)和幂律函数(n 次幂和n 次根变换)。恒等函数是最一般的情况，是输出灰度等于输入灰度的变换，在图 3.3 中包含它仅是为了考虑完整性。

3.2.1 图像反转

使用图3.3 中所示的反转变换，可得到灰度级范围为$[0, L-1]$的一幅图像的反转图像，该反转图像由下式给出：

$$s = L-1-r \tag{3.2-1}$$

使用这种方式反转一幅图像的灰度级，可得到等效的照片底片。这种类型的处理特别适用于增强嵌入图像暗色区域中的白色或灰色细节，特别是当黑色面积在尺寸上占主导地位时。图3.4显示了一个例子。原图像是一幅数字乳房 X 射线照片，其中显示有一小块病变。尽管事实上两幅图在视觉内容上都一样，但应注意，在这种特殊情况下，分析乳房组织时使用反转图像会容易得多。

3.2.2 对数变换

128 ~ 130

图3.3 中的对数变换的通用形式为

$$s = c\log(1+r) \tag{3.2-2}$$

式中 c 是常数，并假设 $r \geqslant 0$。图 3.3 中对数曲线的形状表明，该变换将输入中范围较窄的低灰度值映射为输出中范围较宽的灰度值，或将输入中范围较宽的高灰度值映射为输出中范围较窄的灰度值。我们使用这种类型的变换来扩展图像中的暗像素值，同时压缩更高灰度级的值。反对数变换的作用与此相反。

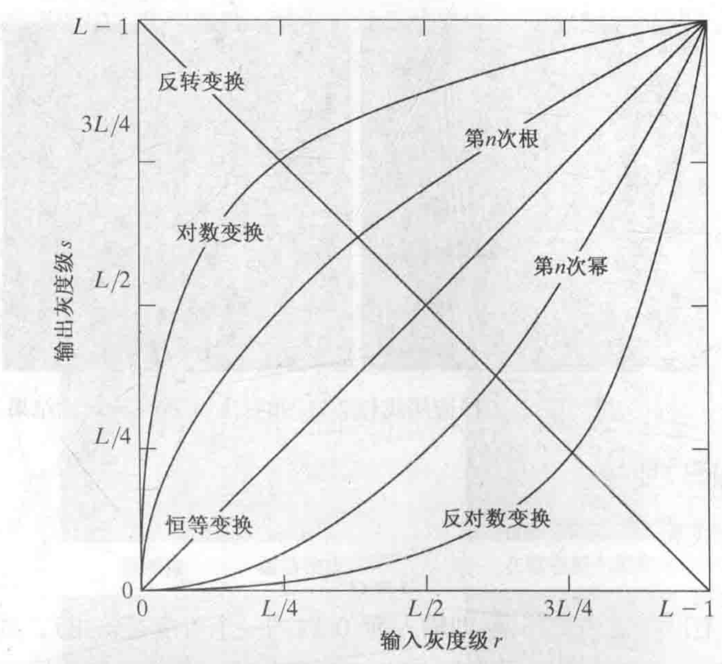


图 3.3　一些基本的灰度变换函数。所有曲线已被缩放到适合显示的范围

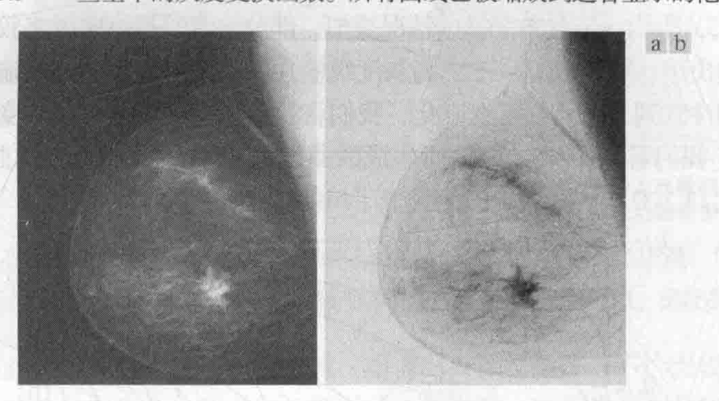


图 3.4　(a) 原始数字乳房 X 射线照片；(b) 使用式 (3.2-1) 给出的反转变换得到的反转图像（图像由 G. E. Medical Systems 公司提供）

具有图 3.3 所示对数函数的一般形状的任何曲线，都能完成图像灰度级的扩展/压缩，但下一节讨论的幂律变换更适用于这一目的。对数函数有一个重要特征，即它压缩像素值变化较大的图像的动态范围。像素值有较大动态范围的一个典型应用说明是傅里叶频谱，这将在第 4 章中讨论。现在，我们只关注图像的频谱特征。通常，频谱值的范围从 0 到 10^6，甚至更高。尽管计算机能毫无问题地处理这一范围的数值，但图像的显示系统通常不能如实地再现如此大范围的灰度值。因而，最终结果是许多重要的灰度细节在典型的傅里叶频谱的显示中丢失了。

作为对数变换的说明，图 3.5(a) 显示了值域为 $0 \sim 1.5\times10^6$ 的傅里叶频谱。当这些值在一个 8 比特系统中被线性地缩放显示时，最亮的像素将支配该显示，频谱中的低值（恰恰是重要的）将损失掉。图3.5(a) 中相对较小的图像区域，鲜明地体现了这种支配性的效果，而作为黑色则观察不到。替代这种显示数值的方法，如果我们先对这些频谱值应用式 (3.3-2)（此时 $c = 1$），那么得到的值域就变为

131 ~ 132

0 ~ 6.2 。图3.5(b)显示了线性缩放这个新值域并在同一个 8 比特显示系统中显示频谱的结果。由这些图片可以看出，与未改进显示的频谱相比，这幅图像中可见细节的丰富程度很明显。我们在关于图像处理的出版物中看到的绝大多数傅里叶频谱，都用这种方式标定过。

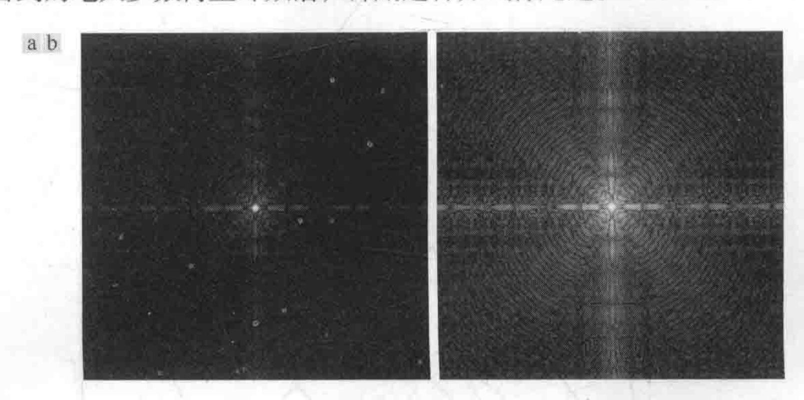


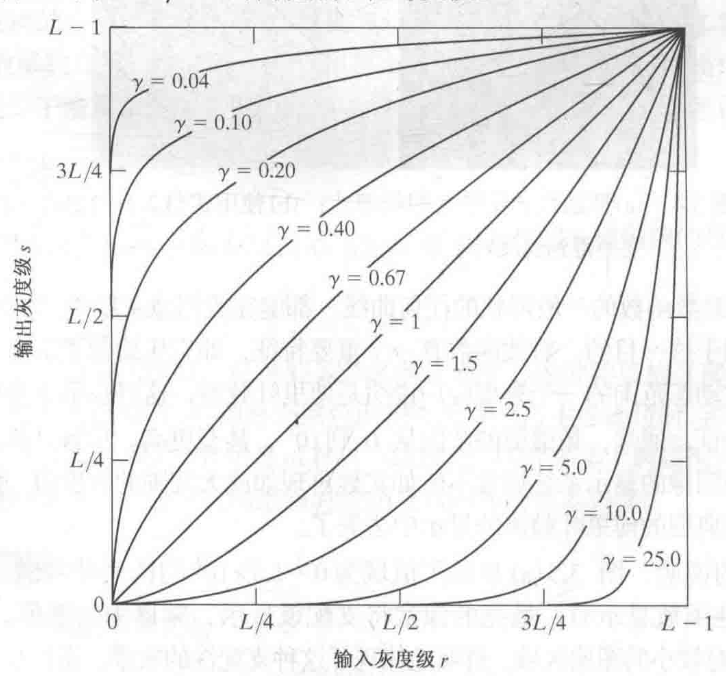
图 3.5 (a)傅里叶频谱；(b)应用式(3.2-2)中的对数变换($c=1$)的结果

3.2.3 幂律(伽马)变换

幂律变换的基本形式为

$$s = cr^{\gamma} \tag{3.2-3}$$

式中 c 和 γ 为正常数。有时考虑到偏移量(即输入为 0 时的一个可度量输出)，式(3.2-3)也写为 $s = c(r+\varepsilon)^{\gamma}$ 。然而，偏移量一般是显示标定问题，因此通常在式(3.2-3)中忽略不计。对于不同的 γ 值，s 与 r 的关系曲线如图3.6所示。与对数变换的情况类似，部分 γ 值的幂律曲线将较窄范围的暗色输入值映射为较宽范围的输出值，或将较宽范围的高灰度级输入值映射为较窄范围的输出值。然而，与对数函数不同的是，我们注意到，随着 γ 值的变化，我们将简单地得到一族可能的变换曲线。正如所预期的那样，在图3.6中，我们看到，$\gamma > 1$ 的值所生成的曲线和 $\gamma < 1$ 的值所生成的曲线的效果完全相反。最后，我们注意到式(3.2-3)在 $c = \gamma = 1$ 时简化成了恒等变换。



132 ~ 133

图 3.6 不同 γ 值的 $s = cr^{\gamma}$ 曲线(所有情况 $c=1$)。所有曲线均已被缩放到适合于显示的范围

用于图像获取、打印和显示的各种设备根据幂律来产生响应。习惯上，幂律方程中的指数称为伽马［故在式(3.2-3)中使用这一符号］。用于校正这些幂律响应现象的处理称为伽马校正。例如，阴极射线管(CRT)设备有一个灰度-电压响应，该响应是一个指数变化范围约为1.8～2.5的幂函数。在图3.6中，用$\gamma = 2.5$时的参考曲线，我们看到，这种显示系统产生的图像往往要比期望的图像暗。图3.7中说明了这一结果。图3.7(a)显示了一幅输入到监视器的简单灰度斜坡(渐变)图像。如期望的那样，监视器表现出了输出比输入暗，如图3.7(b)所示。在这种情况下，伽马校正很简单。我们需要做的只是将图像输入到监视器前进行预处理，即进行$s = r^{1/2.5} = r^{0.4}$变换。结果如图3.7(c)所示。当输入到相同的监视器时，这一伽马校正过的输入产生外观上接近于原图像的输出，如图3.7(d)所示。类似的分析也适用于其他图像设备，如扫描仪和打印机。唯一的不同是随设备而定的伽马值(Poynton[1996])。

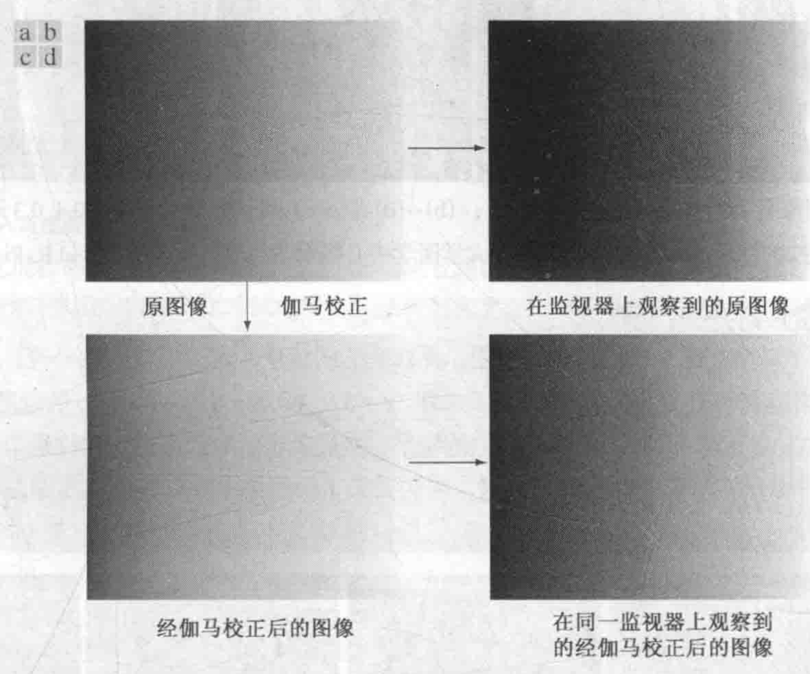


图 3.7　(a)亮度斜坡图像；(b)在伽马值为2.5的模拟监视器上观察到的图像；(c)经伽马校正后的图像；(d)在同一监视器上观察到的经伽马校正后的图像。比较图(d)和图(a)

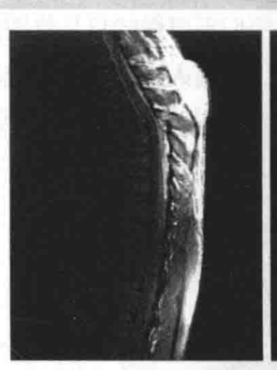
若所关注的是在计算机屏幕上精确显示图像，则伽马校正很重要。不恰当校正的图像看起不是太亮，就是太暗。试图精确再现彩色也需要伽马校正的一些知识，因为改变伽马值不仅会改变亮度，而且会改变彩色图像中红、绿、蓝的比率。随着数字图像在互联网上商业应用的增多，在过去几年里，伽马校正逐渐变得越来越重要。为流行网站创建被几百万人浏览的图像是很平常的事，因为大多数浏览者会有不同的监视器和/或监视器设置。有些计算机系统甚至会内置部分伽马校正。此外，目前的图像标准并不包含创建图像的伽马值，因此问题进一步复杂化了。由于这些限制，在网站中存储图像时，一种合理的方法是用伽马值对图像进行预处理，这个伽马值代表了在开放的市场中，在任意给定时间点，各种型号监视器和计算机系统所期望的“平均值”。

例 3.1　使用幂律变换增强对比度。

除伽马校正外，幂律变换在通用对比度操作中也很有用。图3.8(a)显示了人体胸部以上脊椎骨折错位和脊髓受到影响的一幅核磁共振(MRI)图像。在胸椎垂直中心附近，即图中从上往下约1/4处，骨折显而易见。由于所给图像整体偏暗，因而需要扩展灰度级。这可使用指数为分数的幂律变换来完成。使用式(3.2-3)给出的幂律变换来处理图3.8(a)，得到示于图中的其他几幅图像。对应于图(b)到图(d)的伽马值

133
～
135

分别为 0.6, 0.4 和 0.3(c 始终为 1)。我们注意到，随着伽马值从 0.6 减小到 0.4，更多的细节变得可见。当伽马值进一步减小到 0.3 时，背景中的细节得到了进一步增强，但对比度会降低到图像开始有轻微“冲淡”外观的程度，尤其是在背景中。比较所有的结果，我们看到在对比度和分辨细节方面，最好的增强出现在 $\gamma = 0.4$ 时。在这幅特定图像中，$\gamma = 0.3$ 的值是一个近似极限点，低于该值对比度会下降到令人难以接受的程度。






a b c d

图 3.8 (a)人体脊椎骨折的核磁共振图像(MRI)；(b)~(d)令 $c = 1$ 而 γ 分别等于 0.6, 0.4, 0.3 时应用式(3.2-3)给出的变换的结果(原图像由 Vanderbilt 大学医学中心辐射学与放射学系的 David R. Pickens 博士提供)

例 3.2 幂律变换的另一个说明。

图 3.9(a)显示了一个与图 3.8(a)相反的问题。要处理的图像现在有“冲淡”的外观，表明需要进行灰度级压缩。可以用式(3.2-3)并令 $\gamma > 1$ 来完成这一处理。$\gamma = 3.0, 4.0$ 和 5.0 时，图3.9(a)的处理结果如图3.9(b)到(d)所示。伽马值取 3.0 和 4.0 时，可得到合适的结果，且后者由于有较高的对比度而显示出更令人满意的效果。$\gamma = 5.0$ 时得到的结果中有些地方太暗，因而丢失了一些细节，如左上方主路左侧的暗色区域就是这样的一个例子。

a b c d






图 3.9 (a)航拍图像；(b)~(d)令 $c = 1$ 且 γ 分别等于 3.0, 4.0 和 5.0 时应用式(3.2-3)给出的变换的结果(此例的原图像由 NASA 提供)

3.2.4 分段线性变换函数

前面三节中所讨论方法的一个补充方法是使用分段线性函数。分段线性函数较我们已经讨论过的函数类型的主要优点是，其形式可以任意复杂。事实上，正如我们即将看到的那样，一些重要变换的实际实现可仅由分段函数来明确地表达。分段函数的主要缺点是其技术说明要求用户输入。

对比度拉伸

最简单的分段线性函数之一是对比度拉伸变换。低对比度图像由照明不足、成像传感器动态范围太小、图像获取过程中镜头光圈设置错误引起。对比度拉伸是扩展图像灰度级动态范围的处理，因此可以跨越记录介质和显示装置的全部灰度范围。

135 ~ 137

图3.10(a)显示了一个用于对比度拉伸的典型变换。点(r_1,s_1)和点(r_2,s_2)的位置控制变换函数的形状。若$r_1=s_1$且$r_2=s_2$，则变换为一线性函数，产生没有变化的灰度级。若$r_1=r_2,s_1=0$且$s_2=L-1$，则变换变为阈值处理函数，产生一幅二值图像，如图3.2(b)所示。(r_1,s_1)和(r_2,s_2)的中间值产生输出图像的不同程度的灰度级扩展，从而影响输出图像的对比度。一般情况下，假设$r_1\leqslant r_2$且$s_1\leqslant s_2$，函数是单值的且单调递增的。这一条件保持了灰度级的次序，从而避免了在处理后的图像中产生人为的灰度错误。



图3.10　对比度拉伸：(a)变换函数的形式；(b)低对比度图像；(c)对比度拉伸的结果；(d)阈值处理的结果(原图像由位于堪培拉的澳大利亚国立大学生物科学研究院的Roger Heady博士提供)

图3.10(b)显示了一幅8比特低对比度图像。图3.10(c)显示了对比拉伸后的效果，得到该效果的参数设置如下：$(r_1,s_1)=(r_{\min},0)$且$(r_2,s_2)=(r_{\max},L-1)$，其中$r_{\min}$和$r_{\max}$分别表示图像中的最小灰度级和最大灰度级。因此，变换函数把灰度级由其原范围线性地拉伸至整个范围$[0,L-1]$。最后，图3.10(d)显示了使用前面定义的$(r_1,s_1)=(m,0)$和$(r_2,s_2)=(m,L-1)$的阈值处理函数后的结果，其中，m是图像的平均灰度级。得到这些结果的原图像，是电子显微镜扫描并放大约700倍的花粉图像。

灰度级分层

突出图像中特定灰度范围的亮度通常很重要，其应用包括增强某些特征，如卫星图像中的水体和X射线图像中的缺陷。通常称之为*灰度级分层*的处理可以由许多方法实现，但它们中的大多数是两种基本方法的变形。一种方法是将感兴趣范围内的所有灰度值显示为一个值(譬如“白色”)，而将其他灰度值显示为另一个值(譬如“黑色”)。如图3.11(a)所示，该变换产生了一幅二值图像。第二种方法以图3.11(b)所示的变换为基础，使感兴趣范围的灰度变亮(或变暗)，而保持图像中的其他灰度级不变。

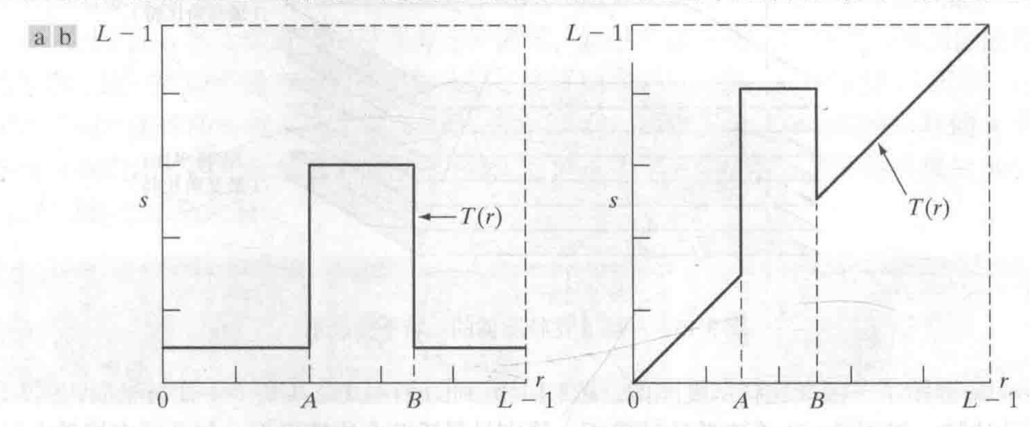


图3.11　(a)这一变换突出了范围$[A,B]$内的灰度，而将所有其他灰度降低到了一个更低的级别；(b)这一变换突出了范围$[A,B]$内的灰度，而保持所有其他灰度级不变

137~138

例 3.3 灰度级分层。

图3.12(a)是一幅靠近肾脏区域的大动脉血管造影照片(见 1.3.2 节对这幅图像的详细解释)。该例的目的是使用灰度级分层来突出主要血管，使其更亮一些，如同注射造影剂的效果那样。图3.12(b)显示了使用图3.11(a)给出的变换的结果，所选范围接近数值的顶端，因为感兴趣的范围比背景亮。该变换的最终结果是血管和肾脏部分显现白色，而其他灰度显示为黑色。这种类型的增强产生了一幅二值图像，它对于研究(用于检测血管堵塞的)造影剂的流动是很有用的。

另一方面，如果我们的兴趣在于研究感兴趣区域的实际灰度值，则可使用图 3.11(b)中的变换。图3.12(c)显示了使用这样一个变换后的结果：近似等于平均灰度的中等灰度区域的灰度范围设置为黑色，而其他灰度保持不变。这里，我们看到主血管和肾脏区域部分的灰度级色调保持不变。当我们的兴趣是测量系列图像中造影剂随时间的流动时，这种结果是很有用的。

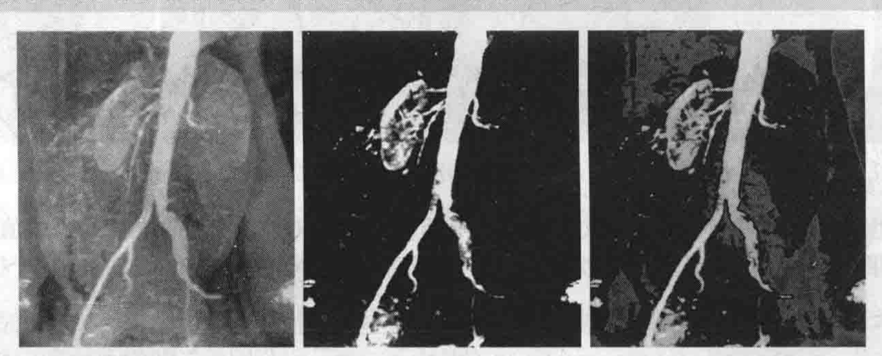


图 3.12 (a)大动脉血管造影照片；(b)使用图 3.11(a)中说明的分层变换的结果，所选的重要灰度范围为灰度级的高端；(c)使用图 3.11(b)中的变换的结果，所选区域置为黑色，以便血管和肾脏区域的灰度保持不变(原图像由密歇根大学医学院的 Thomas R. Gest 博士提供)

比特平面分层

像素是由比特组成的数字。例如，在 256 级灰度图像中，每个像素的灰度是由 8 比特(也就是 1 个字节)组成的。替代突出灰度级范围，我们可突出特定比特来突出整个图像的外观。如图 3.13 说明的那样，一幅 8 比特图像由 8 个 1 比特平面组成，其中平面 1 包含图像中所有像素的最低阶比特，而平面 8 包含图像中所有像素的最高阶比特。

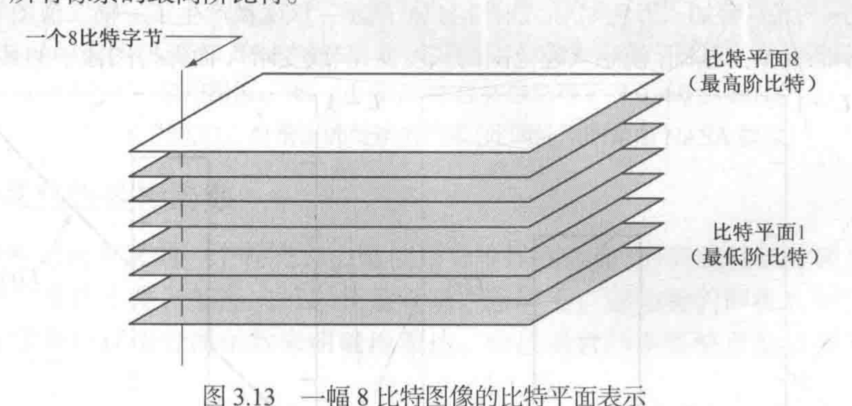


图 3.13 一幅 8 比特图像的比特平面表示

图3.14(a)显示了一幅 8 比特灰度图像，图3.14(b)到图3.14(i)是其 8 个 1 比特平面，图3.14(b)对应于最低阶比特。很明显，4 个高阶比特平面，特别是最后两个比特平面，包含了在视觉上很重要的大多数数据。低阶比特平面在图像中贡献了更精细的灰度细节。原图像有一个灰度为 194 的灰色边界。注意，对应于某些比特平面的灰度边界是黑色(0)，而其他灰度边界是白色(1)。为了解其原因，考虑

139 ~ 140

一个像素，譬如图3.14(a)中最低边界中部的一个像素。从最高阶平面开始，这些比特平面中相应像素的值分别为 1 1 0 0 0 0 1 0，它是十进制数 194 的二进制表示。原图像中任意一个像素的值，可以类似地由这些比特平面中对应的二进制像素值重建。

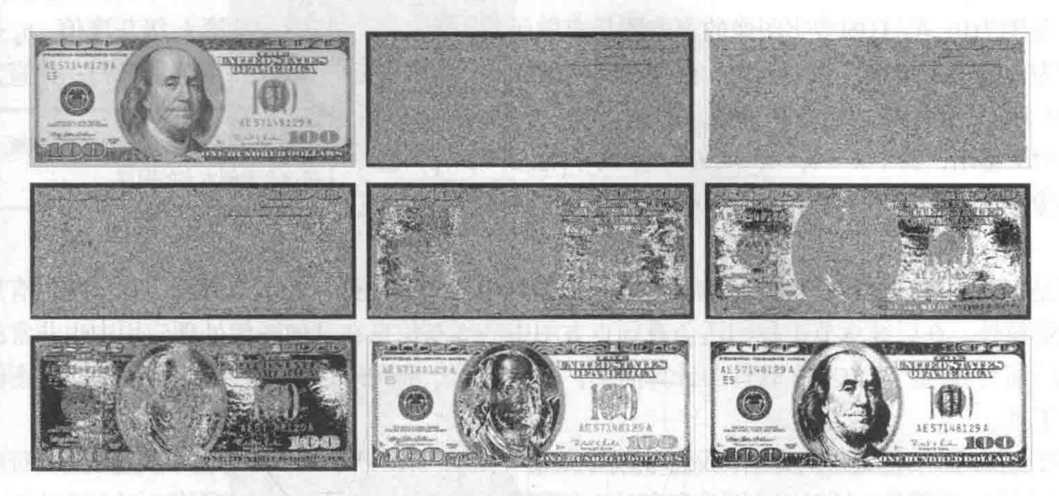


图 3.14　一幅大小为 500×1192 像素的 8 比特灰度图像；(b)~(i) 比特平面 1 到 8，其中比特平面 1 对应于最低阶比特。每个比特平面都是一幅二值图像

根据灰度变换函数，显示一幅 8 比特图像的第 8 个比特平面并不困难，它可用阈值灰度变换函数处理输入图像得到二值图像，该函数将 0~127 之间的所有灰度映射为 0，而将 128~255 之间的所有灰度映射为 1。图 3.14(i) 中的二值图像就是使用这种方式得到的。得到生成其他比特平面的灰度变换函数的方法，将作为一个练习留给读者(见习题 3.4)。

把一幅图像分解为比特平面，对于分析图像中每个比特的相对重要性很有用，这一处理可帮助我们确定用于量化该图像的比特数的充分性。此外，这种类型的分解对图像压缩(第 8 章的主题)也很有用，在图像压缩中，重建一幅图像时所用的平面要比全部平面少。例如，图 3.15(a) 显示了使用比特平面 8 和比特平面 7 重建的一幅图像。重建是使用第 n 个平面的像素乘以常数 2^{n-1} 来完成的。与把第 n 个有意义的二进制比特变换成十进制相比，工作量相同。所用的每个比特平面与一个相应的常数相乘，然后将所有的比特平面相加得到灰度图像。这样，为得到图 3.15(a)，我们用 128 乘以比特平面 8，用 64 乘以比特平面 7，然后把这两个比特平面相加。虽然复原了原图像的主要特征，但重建后的图像显得有些平淡，特别是在背景区域。这并不奇怪，因为两个平面仅能产生 4 个不同的灰度级。把平面 6 加入到重建中可改善这种情况，如图 3.15(b) 所示。注意，该图像的背景有可感知的伪轮廓。这一效应可通过在重建中加入第 5 个平面来明显降低，如图 3.15(c) 所示。在重建中使用更多的平面对图像的外观不会有更大贡献。因而，我们可以得出这样的结论：存储 4 个高阶比特平面将允许我们以可接受的细节来重建原图像。存储这 4 个平面代替原始图像可减少 50% 的存储量(不考虑存储体系结构问题)。





a b c

图 3.15　(a) 使用比特平面 8 和 7 重建的图像；(b) 使用比特平面 8, 7 和 6 重建的图像；(c) 使用比特平面 8, 7, 6 和 5 重建的图像。将图 (c) 与图 3.14(a) 进行比较

140~141

3.3 直方图处理

灰度级范围为$[0, L-1]$的数字图像的直方图是离散函数$h(r_k)=n_k$，其中r_k是第k级灰度值，n_k是图像中灰度为r_k的像素个数。在实践中，经常用乘积MN表示的图像总像素除每个分量来归一化直方图，通常M和N是图像的行数和列数。因此，归一化后的直方图由$p(r_k)=n_k/MN$给出，其中$k=0, 1, \cdots, L-1$。简单地说，$p(r_k)$是灰度级r_k在图像中出现的概率的估计。归一化直方图的所有分量之和应等于1。

> 关于基本概率论的简要回顾，请读者参阅本书的网站。

直方图是多种空间域处理技术的基础。直方图操作可用于图像增强，如本节所示。除了提供有用的图像统计资料外，在后续章节中我们还会看到直方图中的固有信息在其他图像处理应用中也非常有用，如图像压缩与分割。直方图在软件中计算简单，而且有助于商用硬件实现，因此已成为实时图像处理的流行工具。

作为灰度变换的直方图处理的介绍，我们考虑图3.16，它是图3.10所示的以4个基本灰度级为特征的花粉图像：暗图像、亮图像、低对比度图像和高对比度图像。图的右侧显示了与这些图像对应的直方图。每个直方图的水平轴对应于灰度值r_k，垂直轴对应于值$h(r_k)=n_k$或归一化后的值$p(r_k)=n_k/MN$。这样，直方图就可以简单地视为$h(r_k)=n_k$对应于r_k或$p(r_k)=n_k/MN$对应于r_k的图形。

我们注意到，在暗图像中，直方图的分量集中在灰度级的低(暗)端。类似地，亮图像直方图的分量则倾向于灰度级的高端。低对比度图像具有较窄的直方图，且集中于灰度级的中部。对于单色图像，这意味着暗淡，好像灰度被冲淡了一样。最后，我们看到，高对比度图像中直方图的分量覆盖了很宽的灰度级范围，而且像素的分布没有太不均匀，只有少量垂线比其他的高许多。直观上，可以得出这样的结论：若一幅图像的像素倾向于占据整个可能的灰度级并且分布均匀，则该图像会有高对比度的外观并展示灰色调的较大变化。最终效果将是一幅灰度细节丰富且动态范围较大的图像。很快将会证明，仅依靠输入图像直方图中的可用信息就可开发出一个变换函数来自动地实现这种效果。

3.3.1 直方图均衡

考虑连续灰度值，并用变量r表示待处理图像的灰度。通常，我们假设r的取值区间为$[0, L-1]$，且$r=0$表示黑色，$r=L-1$表示白色。在r满足这些条件的情况下，我们将注意力集中在变换形式

$$s=T(r), 0 \leqslant r \leqslant L-1 \tag{3.3-1}$$

上(灰度映射)，对于输入图像中每个具有r值的像素产生一个输出灰度值s。我们假设

(a) $T(r)$在区间$0 \leqslant r \leqslant L-1$上为单调递增函数[①]。
(b) 当$0 \leqslant r \leqslant L-1$时，$0 \leqslant T(r) \leqslant L-1$。

在稍后讨论的一些公式中，我们用反函数

$$r=T^{-1}(s), 0 \leqslant s \leqslant L-1 \tag{3.3-2}$$

在这种情况下，条件(a)改为

(a′) $T(r)$在区间$0 \leqslant r \leqslant L-1$上是一个严格单调递增函数。

142 ~ 144

① 若对于$r_2 > r_1$有$T(r_2) \geq T(r_1)$，则称函数$T(r)$是单调递增的。若对于$r_2 > r_1$有$T(r_2) > T(r_1)$，则称函数$T(r)$是严格单调递增的。类似的定义适用于单调递减函数。

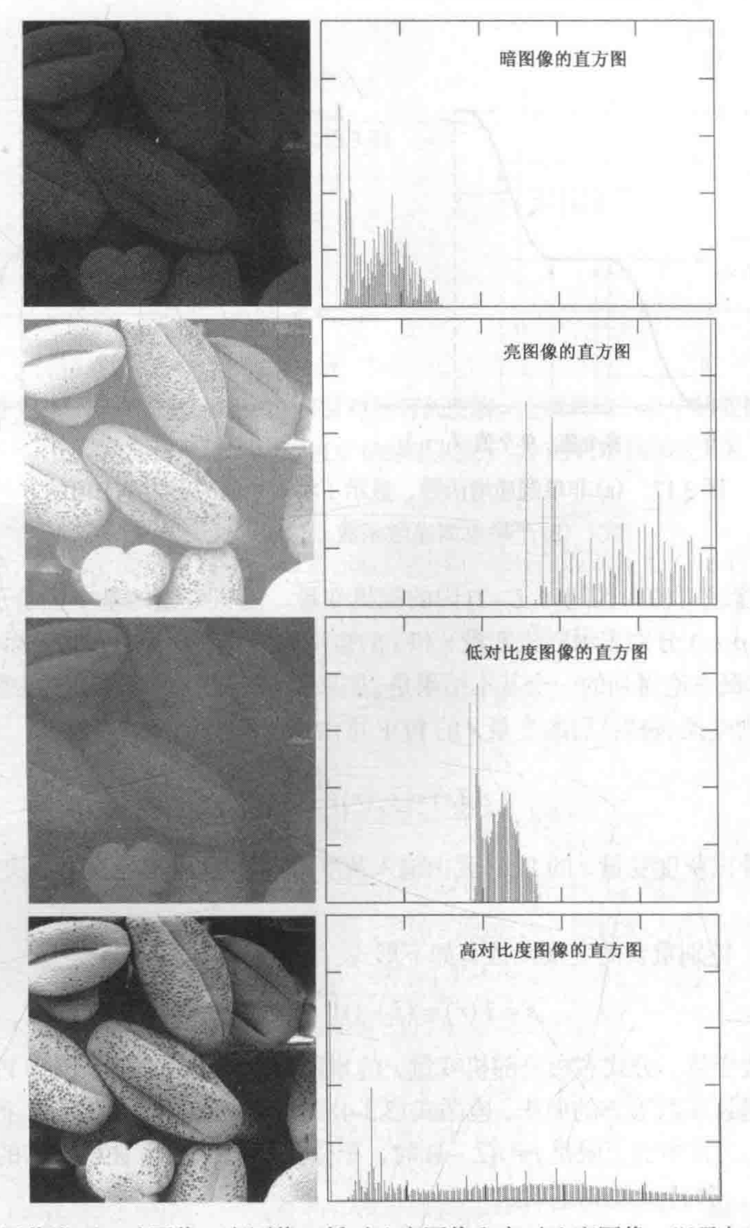


图 3.16　4 种基本的图像类型：暗图像、亮图像、低对比度图像和高对比度图像，以及与它们相对应的直方图

条件(a)中要求 $T(r)$ 为单调递增函数是为了保证输出灰度值不少于相应的输入值，防止灰度反变换时产生人为缺陷。条件(b)保证输出灰度的范围与输入灰度的范围相同。最后，条件(a′)保证从 s 到 r 的反映射是一对一的，防止出现二义性。图3.17(a)显示了满足条件(a)和(b)的一个变换函数。在这里我们看到，多值映射到单值是可能的，并且仍然满足这两个条件。也就是说，单调变换函数执行一对一或多对一映射。当从 r 到 s 映射时，这是很完美的。然而，如果我们想要唯一地从映射的值恢复 r 值(反映射由反向箭头表示)，图3.17(a)就存在一个问题。图3.17(a)中 s_k 的反映射是可能的，但 s_q 的反映射是一个范围值，通常要防止由 s_q 恢复原始 r 值的问题。如图3.17(b)所示，$T(r)$ 要求是严格单调的，以保证反映射是单值的(即两个方向上的映射都是一对一的)。这是我们在本章后面推导一些重要直方图处理技术的理论要求。因为在实践中我们处理的是整数灰度值，必须把所有结果四舍五入为最接近的整数值。因此，不满足严格单调时，我们就要使用寻找最接近匹配整数的方法来解决非唯一反变换的问题。例 3.8 对此给出了说明。

144~145

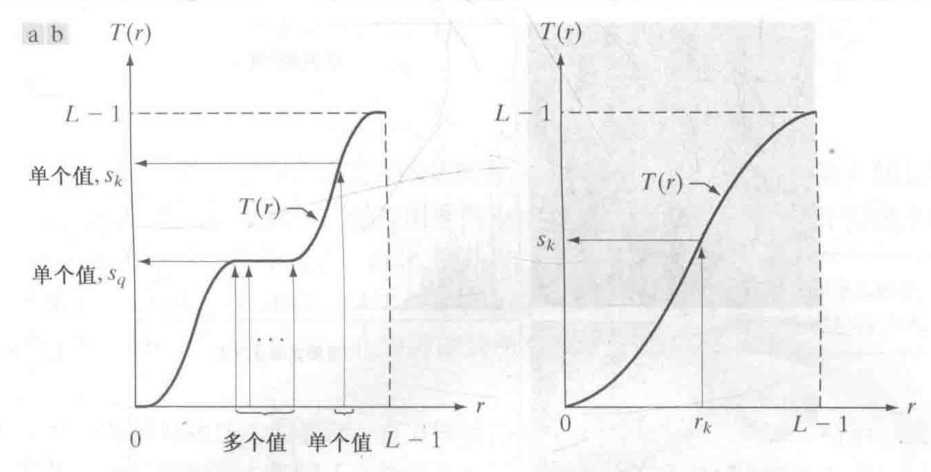


图 3.17　(a)非单调递增函数，显示了将多个值映射为单个值的方式；(b)严格单调递增函数。这是一个双向的一一映射

一幅图像的灰度级可视为区间$[0, L-1]$内的随机变量。随机变量的基本描绘子是其概率密度函数(PDF)。令$p_r(r)$和$p_s(s)$分别表示随机变量r和s的概率密度函数，其中p的下标用于指示p_r和p_s是不同的函数。由基本概率论得到的一个基本结果是，如果$p_r(r)$和$T(r)$已知，且在感兴趣的值域上$T(r)$是连续且可微的，则变换(映射)后的变量s的PDF可由下面的简单公式得到：

$$p_s(s) = p_r(r)\left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| \tag{3.3-3}$$

这样，我们看到，输出灰度变量s的PDF就由输入灰度的PDF和所用变换函数决定［回忆可知r和s由$T(r)$关联起来］。

在图像处理中，特别重要的变换函数有如下形式：

$$s = T(r) = (L-1)\int_0^r p_r(w)\mathrm{d}w \tag{3.3-4}$$

式中，w是积分的假变量。公式右边是随机变量r的累积分布函数(CDF)。因为PDF总为正，回忆可知一个函数的积分是该函数下方的面积，遵循式(3.3-4)的变换函数满足条件(a)，因为函数下的面积不随r的增大而减小。该式中的上限是$r=(L-1)$时，积分值等于1(PDF曲线下方的面积总是1)，所以s的最大值是$(L-1)$，并且条件(b)也满足。

为寻找刚才讨论的相应变换的$p_s(s)$，我们使用式(3.3-3)。由基本积分学中的莱布尼茨准则可知，关于上限的定积分的导数是被积函数在该上限的值，即

$$\frac{\mathrm{d}s}{\mathrm{d}r} = \frac{\mathrm{d}T(r)}{\mathrm{d}r} = (L-1)\frac{\mathrm{d}}{\mathrm{d}r}\left[\int_0^r p_r(w)\mathrm{d}w\right] = (L-1)p_r(r) \tag{3.3-5}$$

把$\mathrm{d}r/\mathrm{d}s$的这个结果代入式(3.3-3)，并记住概率密度值为正，得到

$$p_s(s) = p_r(r)\left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| = p_r(r)\left|\frac{1}{(L-1)p_r(r)}\right| = \frac{1}{L-1},\quad 0 \leqslant s \leqslant L-1 \tag{3.3-6}$$

从该式最后一行中的$p_s(s)$可知，这是一个均匀概率密度函数。简而言之，我们已证明执行式(3.3-4)的灰度变换将得到一个随机变量s，该随机变量由一个均匀PDF表征。特别注意，由该式可知$T(r)$取决于$p_r(r)$，但如式(3.3-6)指出的那样，得到的$p_s(s)$始终是均匀的，它与$p_r(r)$的形式无关。图3.18说明了这些概念。

145 ~ 146

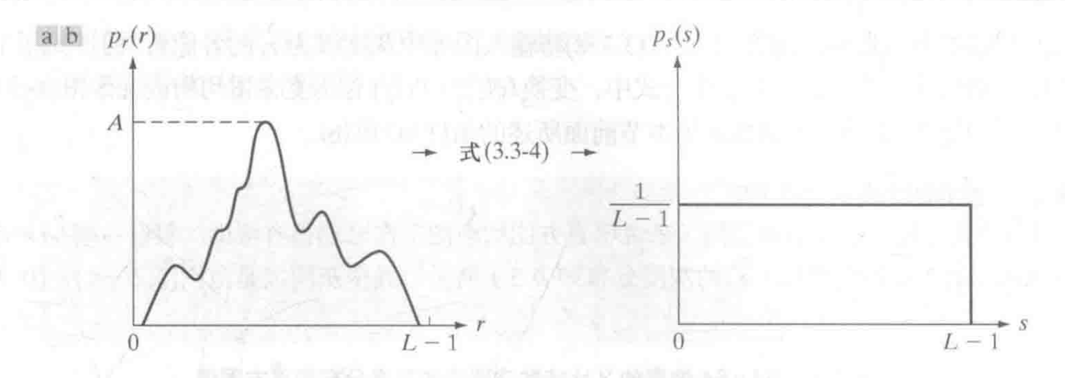


图 3.18 (a)一个任意的 PDF；(b)对所有灰度级 r 应用式(3.3-4)中的变换的结果。具有均匀 PDF 的结果灰度 s 与 r 的 PDF 的形式无关

例 3.4 式(3.3-4)和式(3.3-6)的说明。

为牢记这一概念，考虑下面这个简单的例子。假设图像中的(连续)灰度值有如下 PDF：

$$p_r(r)=\begin{cases}\dfrac{2r}{(L-1)^2}, & 0\leqslant r\leqslant L-1\\ 0, & 其他\end{cases}$$

从式(3.3-4)有

$$s=T(r)=(L-1)\int_0^r p_r(w)\mathrm{d}w=\frac{2}{L-1}\int_0^r w\mathrm{d}w=\frac{r^2}{L-1}$$

假设我们接着使用这个变换得到一幅灰度为 s 的新图像；也就是说，s 值是通过求输入图像的相应灰度值的平方，然后除以 $(L-1)$ 形成的。例如，考虑一幅 $L=10$ 的图像，并假设输入图像中任意位置 (x,y) 处的像素有灰度 $r=3$。则新图像中在该位置的像素是 $s=T(r)=r^2/9=1$。我们可以把 $p_r(r)$ 代入式(3.3-6)，并用 $s=r^2/(L-1)$ 这样的事实，验证新图像中的灰度的 PDF 是均匀的，即

$$p_s(s)=p_r(r)\left|\frac{\mathrm{d}r}{\mathrm{d}s}\right|=\frac{2r}{(L-1)^2}\left|\left[\frac{\mathrm{d}s}{\mathrm{d}r}\right]^{-1}\right|=\frac{2r}{(L-1)^2}\left|\left[\frac{\mathrm{d}}{\mathrm{d}r}\frac{r^2}{L-1}\right]^{-1}\right|=\frac{2r}{(L-1)^2}\left|\frac{(L-1)}{2r}\right|=\frac{1}{L-1}$$

其中最后一步遵循了这样一个事实，即 r 是非负的并且假设 $L>1$。如期望的那样，结果是一个均匀的 PDF。

对于离散值，我们处理其概率(直方图值)与求和来替代处理概率密度函数与积分[①]。如前所述，一幅数字图像中灰度级 r_k 出现的概率近似为

$$p_r(r_k)=\frac{n_k}{MN},\quad k=0,1,2,\cdots,L-1 \tag{3.3-7}$$

式中，MN 是图像的总像素，n_k 是灰度为 r_k 的像素个数，L 是图像中可能灰度级的数量(如对 8 比特图像是 256)。正像本节开始说明的那样，与 r_k 相对的 $p_r(r_k)$ 图形通常称为**直方图**。

式(3.3-4)中变换的离散形式为

$$s_k=T(r_k)=(L-1)\sum_{j=0}^{k}p_r(r_j)=\frac{(L-1)}{MN}\sum_{j=0}^{k}n_j,\quad k=0,1,2,\cdots,L-1 \tag{3.3-8}$$

146 ~ 148

① 先前声明的单调性条件同样适用于离散情形。我们简单地将变量的值限制为离散值。

这样，已处理的图像(即输出图像)通过式(3.3-8)将输入图像中灰度级为 r_k 的各像素映射到输出图像中灰度级为 s_k 的对应像素得到。在这个公式中，变换(映射) $T(r_k)$ 称为直方图均衡或直方图线性变换。不难证明(见习题 3.10)该变换函数满足本节前面所述的条件(a)和(b)。

例 3.5 直方图均衡的简单说明。

在继续之前，通过一个简单的例子来说明直方图均衡的工作原理很有帮助。假设一幅 64×64 像素($MN = 4096$)的 3 比特图像($L = 8$)的灰度分布如表 3.1 所示，其中灰度级是范围$[0, L-1] = [0, 7]$中的整数。

表 3.1 64×64 像素的 3 比特数字图像的灰度分布和直方图值

r_k	n_k	$p_r(r_k) = n_k / MN$
$r_0 = 0$	790	0.19
$r_1 = 1$	1023	0.25
$r_2 = 2$	850	0.21
$r_3 = 3$	656	0.16
$r_4 = 4$	329	0.08
$r_5 = 5$	245	0.06
$r_6 = 6$	122	0.03
$r_7 = 7$	81	0.02

假设图像的直方图如图3.19(a)所示。直方图均衡变换函数的值使用式(3.3-8)得到。例如，

$$s_0 = T(r_0) = 7\sum_{j=0}^{0} p_r(r_j) = 7p_r(r_0) = 1.33$$

类似地，有

$$s_1 = T(r_1) = 7\sum_{j=0}^{1} p_r(r_j) = 7p_r(r_0) + 7p_r(r_1) = 3.08$$

及 $s_2 = 4.55, s_3 = 5.67, s_4 = 6.23, s_5 = 6.65, s_6 = 6.86, s_7 = 7.00$。该变换函数的形状为阶梯状，如图 3.19(b)所示。

在这一点上，s 值一直是分数，因为它们是通过求概率值的和产生的，因此我们要把它们近似为最接近的整数：

$$\begin{aligned} s_0 &= 1.33 \to 1 & s_4 &= 6.23 \to 6 \\ s_1 &= 3.08 \to 3 & s_5 &= 6.65 \to 7 \\ s_2 &= 4.55. \to 5 & s_6 &= 6.86 \to 7 \\ s_3 &= 5.67 \to 6 & s_7 &= 7.00 \to 7 \end{aligned}$$

这些是均衡后的直方图的值。很明显，只有 5 个不同的灰度级。因为 $r_0 = 0$ 被映射为 $s_0 = 1$，在均衡后的图像中有 790 个像素具有该值(见表 3.1)。另外，在图像中有 1023 个像素取 $s_1 = 3$ 这个值，有 850 个像素取 $s_2 = 5$ 这个值。然而，r_3 和 r_4 都被映射为同一个值 6，所以在均衡后的图像中有$(656 + 329) = 985$ 个像素取这个值。类似地，在均衡后的图像中有$(245 + 122 + 81) = 448$ 个像素取 7 这个值。使用 $MN = 4096$ 除这些数，就得到了图 3.19(c)所示的均衡后的直方图。

因为直方图是 PDF 的近似，而且在处理中不允许造成新的灰度级，所以在实际的直方图均衡应用中，很少见到完美平坦的直方图。因此，不像连续的情况，通常不能证明离散直方图均衡会导致均匀的直方图。然而，如我们很快将看到的那样，式(3.3-8)具有展开输入图像直方图的趋势，均衡后的图像的灰度级会跨越更宽的灰度级范围。最终结果是增强了对比度。

148 ~ 149

本节前面讨论了具有覆盖整个灰度范围的灰度值的许多优点。除了产生有这种趋势的灰度之外，刚才推导的方法还有另一个优点，即它是完全“自动的”。换句话说，给定一幅图像，直方图均衡处理

仅涉及执行式(3.3-8)，该式可以直接使用从已知图像提取的信息，而不需要更多的参数说明。我们还注意到实现该技术所要求的计算很简单。

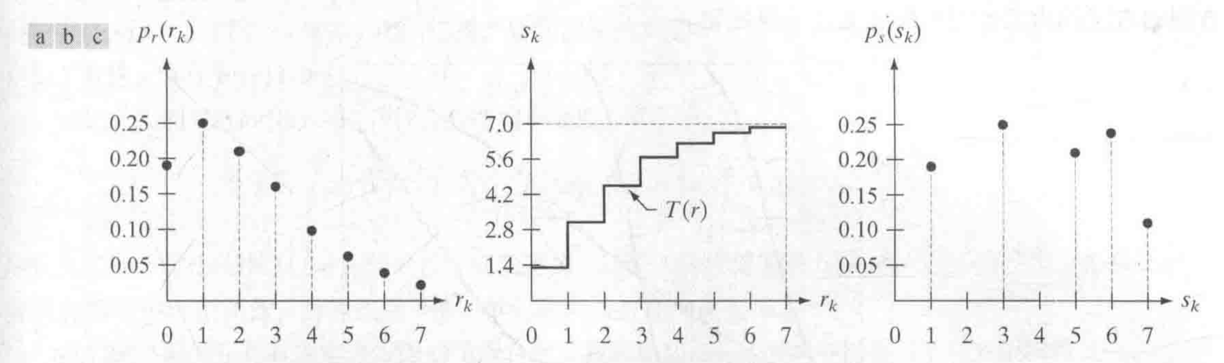


图 3.19　3 比特(8 个灰度级)图像的直方图均衡示例：(a)原始直方图；(b)变换函数；(c)均衡后的直方图

从 s 回到 r 的反变换形式表示为

$$r_k = T^{-1}(s_k), \quad k = 0, 1, 2, \cdots, L-1 \tag{3.3-9}$$

可以证明(见习题 3.10)，只要灰度级 $r_k, k = 0, 1, 2, \cdots, L-1$ 在输入图像中一个也不缺，即图像直方图中没有哪个分量为 0，反变换就满足条件(a')和(b)。尽管反变换不用于直方图均衡，但它在下一节研究的直方图匹配方案中起核心作用。

例 3.6　直方图均衡。

图3.20左边一列显示了来自图3.16的 4 幅图像，中间一列显示了对每幅图像执行直方图均衡后的结果。从上到下的前三种结果表明效果改进明显。如预期的那样，直方图均衡对第四幅图像未产生太大的效果，因为这幅图像的灰度已扩展到了整个灰度级范围。图 3.21 显示了用于产生图 3.20 中均衡后图像的变换函数。这些函数是使用式(3.3-8)产生的。很明显，变换(4)的形状近似为线性形状，它表明输入被映射为近似相等的输出。

图 3.20 中的第三列显示了均衡后的图像的直方图。有趣的是，尽管这些直方图不同，但直方图均衡后的图像本身在视觉上却非常相似。这正是我们所希望的结果，因为左边一列图像之间的差异仅是对比度上的，而不是内容上的。换句话说，因为图像有相同的内容，直方图均衡导致的对比度增强足以补偿图像在视觉上难以区分灰度级的差别。在给出原始图像间的重要对比度差别后，该例说明了直方图均衡作为自适应对比度增强工具的强大作用。

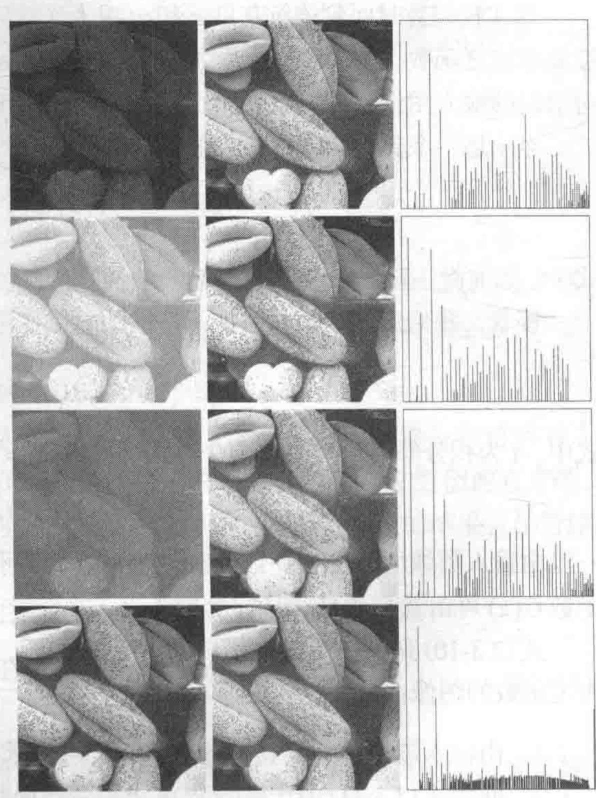


图 3.20　左边一列：来自图 3.16 的图像。中间一列：均衡后的相应图像。右边一列：中间一列图像的直方图

3.3.2　直方图匹配(规定化)

如前所述，直方图均衡能自动地确定变换函数，该函数寻求产生有均匀直方图的输出图像。需要自动增强时，这是一种好方法，因为由这种技术得到的结果可以预知，并且这种方法实现起来

150 ~ 151

也很简单。我们将在本节中指出，对于某些应用，采用均匀直方图的基本增强并不是最好的方法。特别地，有时我们希望处理后的图像具有规定的直方图形状可能更有用。这种用于产生处理后有特殊直方图的方法称为直方图匹配或直方图规定化。

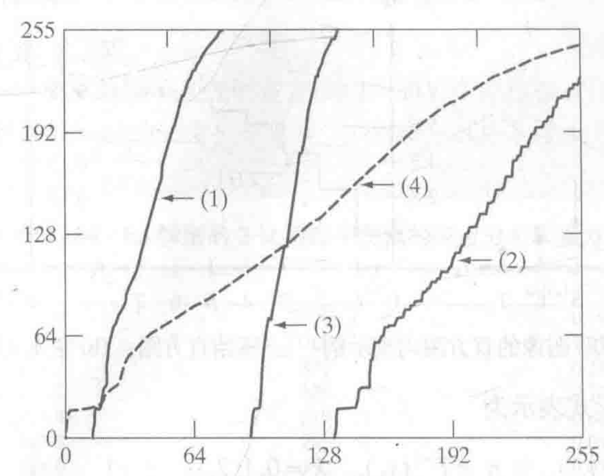


图 3.21 直方图均衡的变换函数。变换(1)到(4)是使用式(3.3-8)由图 3.20 左边一列图像(从上到下)的直方图得到的

现在我们暂时回到连续灰度 r 和 z(视为连续随机变量)，并令 $p_r(r)$ 和 $p_z(z)$ 表示它们所对应的连续概率密度函数。在这种表示法中，r 和 z 分别表示输入图像和输出(已处理)图像的灰度级。我们可以由给定的输入图像估计 $p_r(r)$，而 $p_z(z)$ 是我们希望输出图像所具有的指定概率密度函数。

令 s 是一个具有如下特性的随机变量：

$$s = T(r) = (L-1)\int_0^r p_r(w)\mathrm{d}w \tag{3.3-10}$$

式中，如前面一样，w 为积分假变量。我们发现这个表达式是式(3.3-4)给出的直方图均衡的连续形式。

接着，我们定义一个具有如下特性的随机变量 z：

$$G(z) = (L-1)\int_0^z p_z(t)\mathrm{d}t = s \tag{3.3-11}$$

式中，t 为积分假变量。由这两个等式可得 $G(z) = T(r)$，因此 z 必须满足下列条件：

$$z = G^{-1}[T(r)] = G^{-1}(s) \tag{3.3-12}$$

一旦由输入图像估计出 $p_r(r)$，变换函数 $T(r)$ 就可由式(3.3-10)得到。类似地，因为 $p_z(z)$ 已知，变换函数 $G(z)$ 可由式(3.3-11)得到。

式(3.3-10)到式(3.3-12)表明，使用下列步骤，可由一幅给定图像得到一幅其灰度级具有指定概率密度函数的图像：

1. 由输入图像得到 $p_r(r)$，并由式(3.3-10)求得 s 的值。
2. 使用式(3.3-11)中指定的 PDF 求得变换函数 $G(z)$。
3. 求得反变换函数 $z = G^{-1}(s)$；因为 z 是由 s 得到的，所以该处理是 s 到 z 的映射，而后者正是我们期望的值。
4. 首先用式(3.3-10)对输入图像进行均衡得到输出图像；该图像的像素值是 s 值。对均衡后的图像中具有 s 值的每个像素执行反映射 $z = G^{-1}(s)$，得到输出图像中的相应像素。当所有像素都处理完后，输出图像的 PDF 将等于指定的 PDF。

150 ~ 153

例 3.7　直方图规定化。

假设采用连续灰度值，并假设一幅图像的灰度 PDF 为 $p_r(r)=2r/(L-1)^2$，$0\leqslant r\leqslant(L-1)$，对于其他 r 值有 $p_r(r)=0$。寻找一个变换函数，使得产生的图像的灰度 PDF 是 $p_z(z)=3z^2/(L-1)^3$，$0\leqslant z\leqslant(L-1)$，而对于其他 z 值有 $p_z(z)=0$。

首先，我们对区间 $[0,L-1]$ 寻找直方图均衡变换：

$$s=T(r)=(L-1)\int_0^r p_r(w)\mathrm{d}w=\frac{2}{(L-1)}\int_0^r w\mathrm{d}w=\frac{r^2}{(L-1)}$$

由定义可知，对于范围 $[0,L-1]$ 外的值，该变换为 0。输入灰度值的平方除以 $(L-1)^2$ 将产生一幅灰度为 s 并具有均匀 PDF 的图像，因为这是早些时候讨论的直方图均衡变换。

我们感兴趣的是具有规定直方图的图像，因此我们在 $[0,L-1]$ 区间寻找下一个直方图均衡变换：

$$G(z)=(L-1)\int_0^z p_z(w)\mathrm{d}w=\frac{3}{(L-1)^2}\int_0^z w^2\mathrm{d}w=\frac{z^3}{(L-1)^2}$$

由定义可知在该区间外该函数也为 0。最后，我们要求 $G(z)=s$，但 $G(z)=z^3/(L-1)^2$；因此 $z^3/(L-1)^2=s$，并且有

$$z=[(L-1)^2 s]^{1/3}$$

如果用 $(L-1)^2$ 乘以直方图均衡过的每个像素，取该乘积的 1/3 次方，结果将是一幅我们所期望的图像，该图像在区间 $[0,L-1]$ 内 z 的灰度的 PDF 为 $p_z(z)=3z^2/(L-1)^3$。

因为 $s=r^2/(L-1)$，我们可以直接由输入图像的灰度 r 生成 z：

$$z=[(L-1)^2 s]^{1/3}=\left[(L-1)^2\frac{r^2}{(L-1)}\right]^{1/3}=\left[(L-1)r^2\right]^{1/3}$$

这样，原图像中每个像素值的平方与 $(L-1)$ 相乘，然后再取该乘积的 1/3 次方，将得到其灰度级 z 具有规定PDF的图像。我们看到，均衡输入图像的中间一步可以跳过；我们需要得到将 r 映射为 s 的变换函数 $T(r)$。然后，这两步可以合并为从 r 到 z 的一步变换。

正如前面的例子所示，直方图规定化在原理上很简单。在实践中，常见的困难是寻找 $T(r)$ 和 G^{-1} 的有意义表达式。所幸的是，在处理离散量时，问题可大大简化。这里仅希望得到一个近似的直方图，所以付出的代价与直方图均衡一样。然而，不管如何，我们都可以得到一些非常有用的结果，尽管结果是粗糙的近似。

式(3.3-10)的离散形式是式(3.3-8)中的直方图均衡变换，为方便起见，我们重写如下：

$$s_k=T(r_k)=(L-1)\sum_{j=0}^{k}p_r(r_j)=\frac{(L-1)}{MN}\sum_{j=0}^{k}n_j,\quad k=0,1,2,\cdots,L-1 \tag{3.3-13}$$

与前面一样，式中 MN 是图像的像素总数，n_j 是具有灰度值 r_j 的像素数，L 是图像中可能的灰度级数。类似地，给出一个规定的 s_k 值，式(3.3-11)的离散形式涉及计算变换函数

$$G(z_q)=(L-1)\sum_{i=0}^{q}p_z(z_i) \tag{3.3-14}$$

对一个 q 值，有

$$G(z_q)=s_k \tag{3.3-15}$$

153 ~ 154

式中，$p_z(z_i)$是规定直方图的第i个值。与前面一样，我们用反变换找到期望的值z_q：

$$z_q = G^{-1}(s_k) \tag{3.3-16}$$

换句话说，该操作对每个s值给出一个z值；这样，就形成了从s到z的一个映射。

实践中，我们不需要计算G的反变换。因为我们处理的灰度级是整数(如8比特图像的灰度级是从0到255)，使用式(3.3-14)计算$q=0,1,2,\cdots,L-1$时的所有可能G值很简单。标定这些值，并四舍五入为区间$[0,L-1]$内的最接近整数。将这些值存储在一个表中。给定一个特殊的s_k值后，我们可以查找存储在表中的最匹配的值。例如，如果在表中第64个输入接近s_k，则$q=63$(注意，我们是从0开始计数的)，故z_{63}是式(3.3-15)的最优解。这样，给定的s_k值就与z_{63}关联在一起(即规定的s_k值将映射为z_{63})。因为z项是规定直方图$p_z(z)$所用的灰度，满足$z_0=0,z_1=1,\cdots,z_{L-1}=L-1$，因此，$z_{63}$的灰度值为63。重复该过程，我们可找到每个$s_k$值到$z_q$值的映射，它们是式(3.3-15)的最接近的解。这些映射也是直方图规定化问题的解。

回忆s_k是直方图均衡后的图像的值，我们可以总结直方图规定化过程如下：

1. 计算给定图像的直方图$p_r(r)$，并用它寻找式(3.3-13)的直方图均衡变换。把s_k四舍五入为范围$[0,L-1]$内的整数。
2. 用式(3.3-14)对$q=0,1,2,\cdots,L-1$计算变换函数G的所有值，其中$p_z(z_i)$是规定直方图的值。把G的值四舍五入为范围$[0,L-1]$内的整数。将G的值存储在一个表中。
3. 对每个值$s_k,k=0,1,2,\cdots,L-1$，使用步骤2存储的G值寻找相应的z_q值，以使$G(z_q)$最接近s_k，并存储这些从s到z的映射。当满足给定s_k的z_q值多于一个时(即映射不唯一时)，按惯例选择最小的值。
4. 先对输入图像进行均衡，然后使用步骤3找到的映射把该图像中每个均衡后的像素值s_k映射为直方图规定化后的图像中的相应z_q值，形成直方图规定化后的图像。如连续情况那样，均衡输入图像的中间步骤是概念上的。合并两个变换函数T和G^{-1}可跳过这一步，如例3.8所示。

如前面提到的那样，对于满足条件(a')和(b)的G^{-1}，G必须是严格单调的，根据式(3.3-14)，它意味着规定直方图的任何$p_z(z_i)$值都不能为零(见习题3.10)。当工作在离散数值的情况时，该条件可能不满足的事实并不是一个严重的实现问题，如上面的步骤3中指出的那样。下面的例子用数字说明了这一情况。

例3.8 直方图规定化的一个简单例子。

再次考虑例3.5中大小为64×64的假设图像，其直方图重复显示在图3.22(a)中。希望变换该直方图，以便使其具有表3.2中第二列规定的值。图3.22(b)显示了该直方图的大概形状。

表3.2 规定直方图和实际直方图(第三列的值来自例3.8中的计算结果)

z_q	规定的$p_z(z_q)$	实际的$p_z(z_k)$
$z_0=0$	0.00	0.00
$z_1=1$	0.00	0.00
$z_2=2$	0.00	0.00
$z_3=3$	0.15	0.19
$z_4=4$	0.20	0.25
$z_5=5$	0.30	0.21
$z_6=6$	0.20	0.24
$z_7=7$	0.15	0.11

154~155

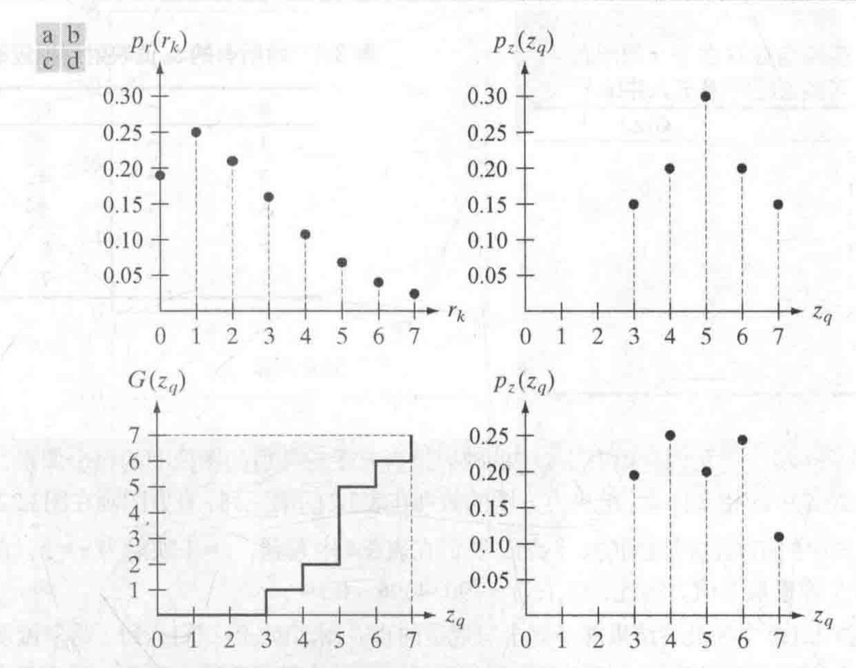


图 3.22　(a)一幅 3 比特图像的直方图；(b)规定的直方图；(c)从规定的直方图得到的变换函数；(d)执行直方图规定化的结果。请比较图(b)和图(d)

过程的第一步是得到标定直方图均衡后的值，就像在例 3.5 中那样：

$$s_0 = 1 \quad s_2 = 5 \quad s_4 = 6 \quad s_6 = 7$$
$$s_1 = 3 \quad s_3 = 6 \quad s_5 = 7 \quad s_7 = 7$$

接着，我们使用式(3.3-14)来计算变换函数 G 的所有值：

$$G(z_0) = 7\sum_{j=0}^{0} p_z(z_j) = 0.00$$

类似地，有

$$G(z_1) = 7\sum_{j=0}^{1} p_z(z_j) = 7\left[p(z_0) + p(z_1)\right] = 0.00$$

和

$$G(z_2) = 0.00, G(z_4) = 2.45, G(z_6) = 5.95, G(z_3) = 1.05, G(z_5) = 4.55, G(z_7) = 7.00\text{。}$$

如例 3.5 那样，要把这些分数值转换为有效区间[0, 7]内的整数。结果是

$$G(z_0) = 0.00 \to 0 \quad G(z_4) = 2.45 \to 2$$
$$G(z_1) = 0.00 \to 0 \quad G(z_5) = 4.55 \to 5$$
$$G(z_2) = 0.00 \to 0 \quad G(z_6) = 5.95 \to 6$$
$$G(z_3) = 1.05 \to 1 \quad G(z_7) = 7.00 \to 7$$

这些结果总结在表3.3中，变换函数画在图3.22(c)中。观察发现 G 并不是严格单调的，所以违反了条件(a′)。因此，我们使用算法步骤 3 中描述的方法来处理这种情况。

在过程的第三步中，我们找到 z_q 的最小值，所以值 $G(z_q)$ 最接近 s_k。对每个 s_k 值这样做，以产生从 s 到 z 所需要的映射。例如，$s_0 = 1$，我们看到 $G(z_3) = 1$，在这种情况下，这是完美的匹配，因此我们对应有 $s_0 \to z_3$。也就是说，直方图均衡后的图像中的每个值为 1 的像素，映射为直方图规定化后的图像中(相应位置上)的值为 3 的像素。继续采用这一方法，我们得到表 3.4 中的映射。

155~157

表 3.3 变换函数 G 关于 z 的所有可能值(已四舍五入并排序)

z_q	$G(z_q)$
$z_0=0$	0
$z_1=1$	0
$z_2=2$	0
$z_3=3$	1
$z_4=4$	2
$z_5=5$	5
$z_6=6$	6
$z_7=7$	7

表 3.4 将所有的 s_k 值映射到相应的 z_q 值

s_k	→	z_q
1	→	3
3	→	4
5	→	5
6	→	6
7	→	7

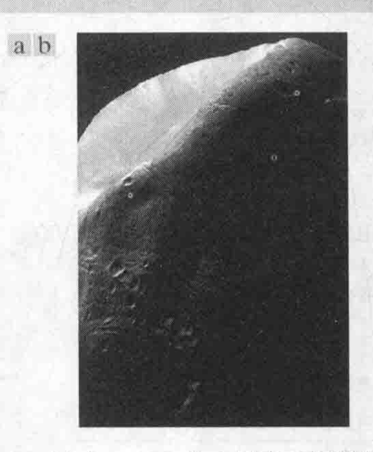
在过程的最后一步中，我们使用表 3.4 中的映射把直方图均衡后的图像中的每个像素，映射为新创建的直方图规定化图像中的相应像素。结果直方图的值列在表3.2 的第三列，直方图画在图3.22(d)中。$p_z(z_q)$ 的值是使用例 3.5 中的相同过程得到的。例如，我们在表 3.4 中看到，$s=1$ 映射为 $z=3$，在直方图均衡后的图像中有 790 个像素取 1 值。因此，$p_z(z_3)=790/4096=0.19$。

虽然示于图3.22(d)中的最终结果并不完全与规定的直方图匹配，但我们达到了将灰度明确地移向灰度级高端的目的。如前面提到的那样，得到直方图均衡后图像的中间步骤对于解释该过程是有帮助的，但它并不是必需的。相反，我们可以在表中的第三列列出从 r 到 s 的映射和从 s 到 z 的映射。然后，使用这些映射把原始像素直接映射到直方图规定化后的图像的像素。

例 3.9 直方图均衡与直方图匹配的比较。

图3.23(a)显示了火卫一的图像，该图像由 NASA 的火星表面探测器拍摄。图3.23(b)显示了图3.23(a)的直方图。图像的大部分是暗色区域，产生了像素集中于灰度级暗端的直方图的特点。乍看之下，会得到直方图均衡化是增强该图像的一个好办法的结论，以便使暗区域的细节更清楚。下面的讨论将证明事实并非如此。

a b



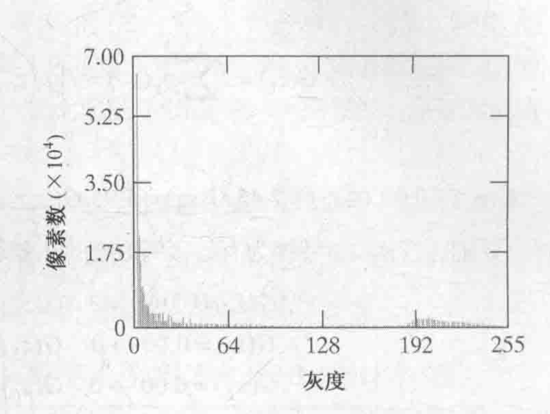


图 3.23 (a)由 NASA 的火星表面探测器拍摄的火卫一图像；(b)直方图(原图像由 NASA 提供)

图3.24(a)显示了由图3.23(b)所示直方图得到的直方图均衡变换[式(3.3-8)或式(3.3-13)]。这个变换函数最相关的特征是，它从灰度 0 快速上升到接近 190。这是因为输入直方图中有大量的像素灰度接近 0。当这种变换应用于输入图像的灰度，以获得直方图的均衡效果时，最终结果是把非常窄的暗像素区间映射到输出图像灰度级的高端。因为输入图像的大量像素在此区间有精确的灰度值，所以我们希望结果是具有明亮且“冲淡”外观的图像。如图3.24(b)所示，情况确实如此。该图像的直方图示于图 3.24(c)。注意，所有灰度级基本上都偏向了灰度范围的上半部。

157 ~ 158

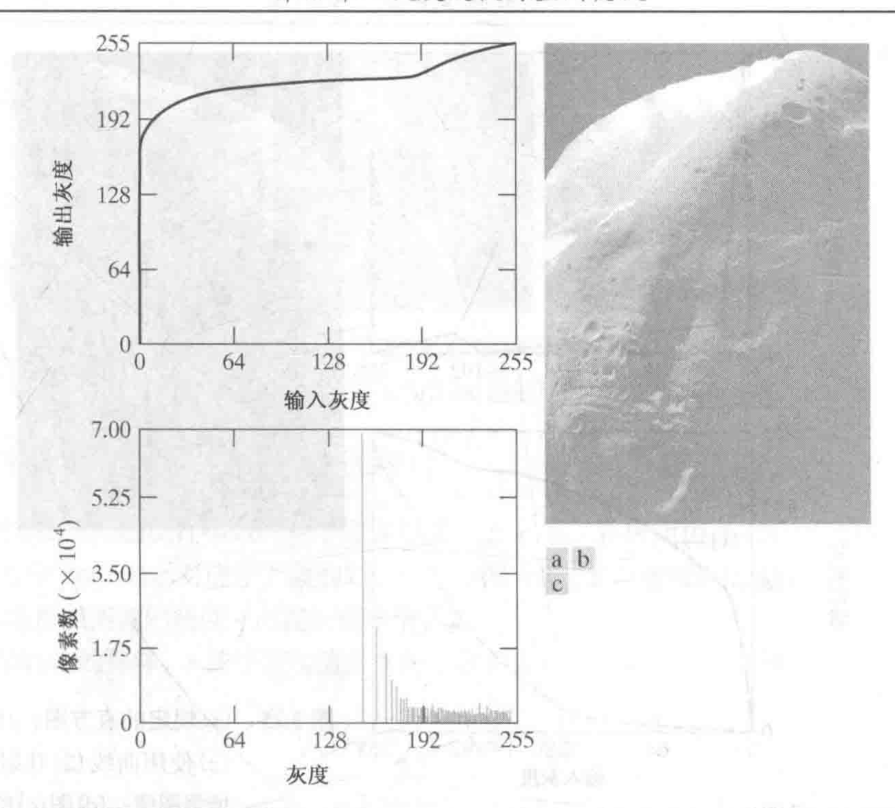


图 3.24　(a) 直方图均衡的变换函数；(b) 直方图均衡后的图像(注意“冲淡”的外观)；(c) 图(b)的直方图

由于图 3.24(a) 中变换函数的问题是由原图像中大量灰度接近于 0 的像素引起的，一种合理的方法是修改图像的直方图，使其不再具有这一属性。图 3.25(a) 显示了一个人为规定的函数，该函数保持了原有直方图的基本形状，但在图像灰度暗区的灰度值却有更平滑的过渡。将这个函数取样为 256 个等间隔的离散值，就生成了所需的规定直方图。使用式 (3.3-14) 从该直方图得到的变换函数 $G(z)$ 在图 3.25(b) 中标注为变换 (1)。类似地，由式 (3.3-16)(使用前面讨论的逐步过程得到) 导出的反变换函数 $G^{-1}(s)$ 在图 3.25(b) 中标注为变换 (2)。图 3.25(c) 中增强后的图像是把变换 (2) 用于图 3.24(b) 中直方图均衡化后的图像的像素得到的。比较这两幅图像可以看出，在使用直方图均衡得到的结果的基础上，直方图规定化图像的改进是很明显的。注意，原始直方图的细小改变会使图像的外观得到显著改进。图 3.25(d) 显示了图 3.25(c) 的直方图。该直方图最容易辨别的特征是，其低端已右移到了灰度级的较亮区域(但并不过分)，这正是我们所希望的。

到目前为止，虽然我们对直方图规定化已经很清楚，但在结束这一节之前，我们要强调直方图规定化在大多数时候都是试凑过程。正像前边的例子中我们所做的那样，有用的指导方式是得用手边的问题进行学习。但是，有时可能会遇到这样的情况，即定义“平均”直方图的外观，并把它作为规定的直方图。在这种情况下，直方图规定化就变成了一种直接的处理。然而，一般来说，并没有规定直方图的规则，对于任何给定的增强任务都必须借助于实际分析。

3.3.3　局部直方图处理

前两节讨论的直方图处理方法是全局性的，在某种意义上，像素被基于整幅图像的灰度分布的变换函数修改。虽然这种全局方法适用于整个图像的增强，但存在增强图像中小区域的细节的需要。在这些区域中，一些像素的影响在全局变换的计算中可能被忽略，因为全局变换没有必要保证期望的局部增强。解决方法是以图像中每个像素的邻域中的灰度分布为基础设计变换函数。

158 ~ 161

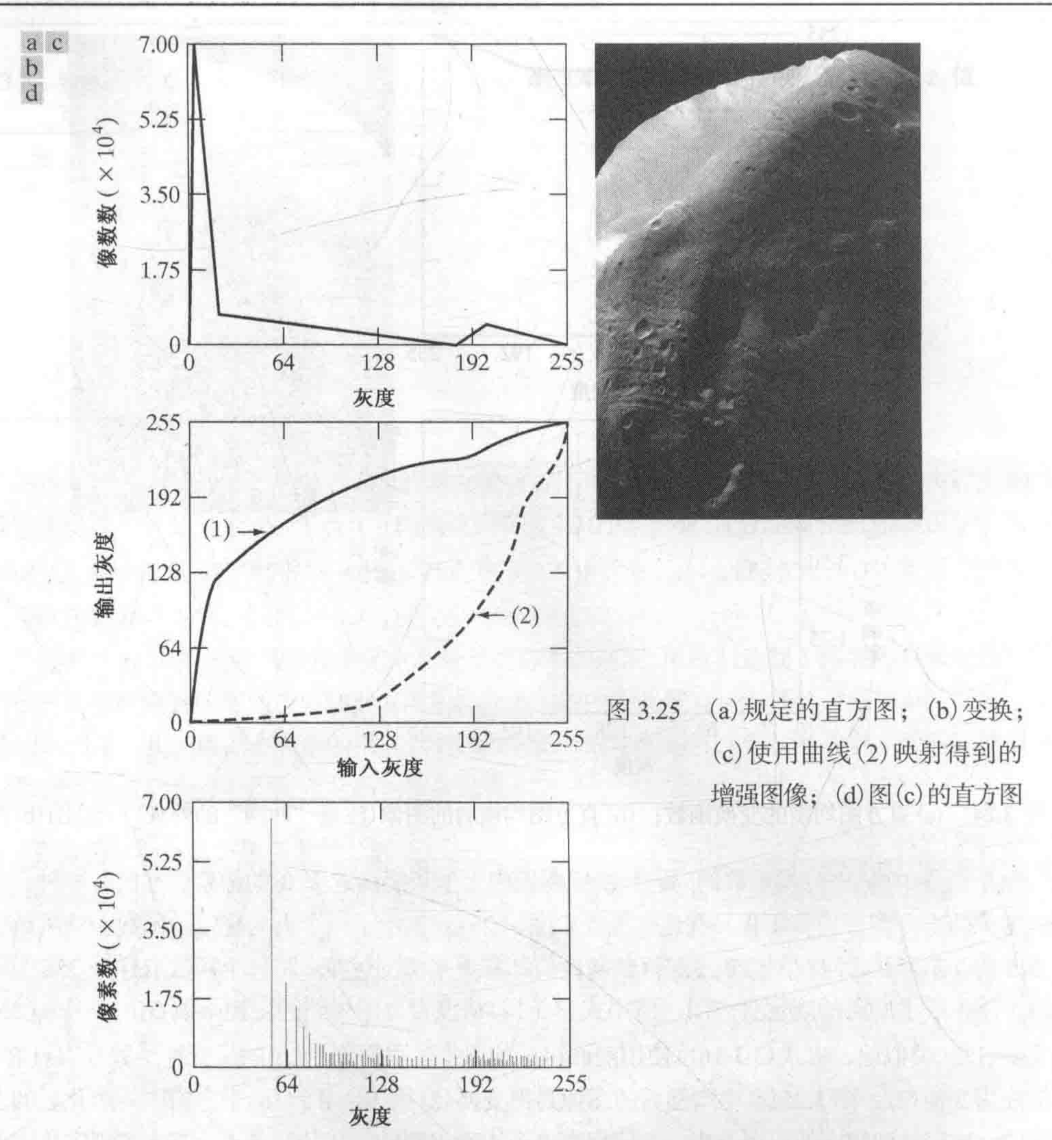


图 3.25 (a)规定的直方图；(b)变换；(c)使用曲线(2)映射得到的增强图像；(d)图(c)的直方图

前面描述的直方图处理技术很容易用于局部增强。过程是定义一个邻域，并把该区域的中心从一个像素移至另一个像素。在每个位置，计算邻域中的点的直方图，并且得到的不是直方图均衡化，就是规定化变换函数。这个函数最终用于映射邻域中心像素的灰度。然后，邻域的中心被移至一个相邻像素位置，并重复该过程。对邻域进行逐像素平移时，由于只有邻域中的一行或一列改变，所以可在每一步移动中，以新数据更新前一位置得到的直方图(见习题 3.12)。这种方法与区域每移动一个像素位置就计算邻域中所有像素的直方图相比优点明显。有时用于减少计算量的另一种方法是，使用非重叠区域，但这种方法通常会产生我们不希望的“棋盘”效应。

例 3.10 局部直方图均衡。

图3.26(a)显示了一幅大小为512×512 的 8 比特图像，初看之下，该图像在灰色背景上包含 5 个黑色方块。图像有轻微的噪声，但感觉不到。图3.26(b)显示了全局直方图均衡的结果。用直方图均衡平滑噪声区域时，情况经常如此。该图像显示噪声明显增强。然而，除了噪声之外，图3.26(b)与原图像相比并未显示新的重要细节，只是不明显地给出了左上方和右下方方块中包含有一个物体。图3.26(c)是使用大小为3×3 的邻域的局部直方图均衡得到的。这里，我们看到了包含在暗色方块中的物体。这些物体的灰度值太接近大方块的灰度，而且它们太小，因此全局直方图均衡化不足以影响这些细节的显示。

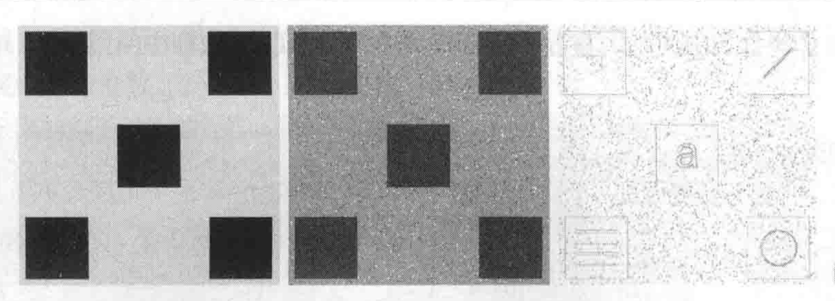


图 3.26　(a)原图像；(b)全局直方图均衡的结果；(c)使用大小为 3×3 的邻域对图(a)应用局部直方图均衡的结果

3.3.4　在图像增强中使用直方图统计

直接从直方图获得的统计参数可用于图像增强。令 r 表示在区间 $[0, L-1]$ 上代表灰度值的一个离散随机变量，并令 $p(r_i)$ 表示对应于 r_i 值的归一化直方图分量。如前面指出的那样，我们可以把 $p(r_i)$ 视为得到直方图的那幅图像的灰度 r_i 出现的概率估计。

像 2.6.8 节讨论的那样，r 关于其均值的 n 阶矩定义为

$$\mu_n(r) = \sum_{i=0}^{L-1} (r_i - m)^n p(r_i) \tag{3.3-17}$$

式中，m 是 r 的均值(平均灰度，即图像中像素的平均灰度)：

$$m = \sum_{i=0}^{L-1} r_i p(r_i) \tag{3.3-18}$$

我们遵守使用 m 来表示均值的约定。请不要将它与用于表示一个大小为 $m \times n$ 的邻域的行数的相同符号混淆，因为在表示邻域的行数时，我们也遵守了符号表示约定。

二阶矩特别重要：

$$\mu_2(r) = \sum_{i=0}^{L-1} (r_i - m)^2 p(r_i) \tag{3.3-19}$$

我们将该表达式称为*灰度方差*，通常用σ^2 表示(回忆可知标准差是方差的平方根)。均值是平均灰度的度量，方差(或标准差)是图像对比度的度量。显然，一旦从给定的图像得到直方图，用前面的表达式就很容易计算所有的矩。

在仅处理均值和方差时，实际上通常直接从取样值来估计它们，而不必计算直方图。近似地，这些估计称为*取样均值*和*取样方差*。根据基本的统计学，它们可由下面熟悉的形式给出：

$$m = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \tag{3.3-20}$$

和

$$\sigma^2 = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[f(x, y) - m \right]^2 \tag{3.3-21}$$

式中，$x = 0, 1, 2, \cdots, M-1$，$y = 0, 1, 2, \cdots, N-1$。换句话说，众所周知，一幅图像的平均灰度可以通过求所有像素的灰度值之和并用图像中的像素总数去除得到。类似的解释适用于式(3.3-21)。如下面的例子所示，使用这两个公式得到的结果等同于使用式(3.3-18)和式(3.3-19)

式(3.3-21)中的分母有时写为 $MN-1$ 而非 MN。目的是获得方差的无偏估计。但我们对式(3.3.21)和式(3.3-19)［在后一公式中的直方图是由式(3.3-21)使用的同一幅图像计算得到的时］一致更感兴趣。因此，我们要求 MN 项。对于实际大小的任何图像，差别可以忽略不计。

162~163

得到的结果，前提是这些公式中使用的直方图是由式(3.3-20)和式(3.3-21)中使用的同一幅图像计算得到的。

例 3.11　直方图统计计算。

在继续讨论之前，我们有必要通过一个简单的数字例子来巩固概念。考虑下面的大小为 5×5 的 2 比特图像：

$$\begin{matrix} 0 & 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 0 & 1 \\ 3 & 3 & 2 & 2 & 0 \\ 2 & 3 & 1 & 0 & 0 \\ 1 & 1 & 3 & 2 & 2 \end{matrix}$$

像素由 2 比特表示；因此，$L = 4$ 且灰度级在[0, 3]范围内。总像素数是 25，因此直方图分量为

$$p(r_0) = \frac{6}{25} = 0.24, \quad p(r_1) = \frac{7}{25} = 0.28$$

$$p(r_2) = \frac{7}{25} = 0.28, \quad p(r_3) = \frac{5}{25} = 0.20$$

式中，$p(r_i)$ 中的分子是具有灰度级 r_i 的图像中的像素数。我们可用式(3.3-18)计算图像中灰度级的平均值：

$$m = \sum_{i=0}^{3} r_i p(r_i) = (0)(0.24) + (1)(0.28) + (2)(0.28) + (3)(0.20) = 1.44$$

令 $f(x, y)$ 表示前面的5×5阵列，使用式(3.3-20) 得到

$$m = \frac{1}{25}\sum_{x=0}^{4}\sum_{y=0}^{4} f(x, y) = 1.44$$

如期望的那样，结果一致。类似地，使用式(3.3-19)或式(3.3-21)得到的方差相同(均为 1.1264)。

我们考虑均值和方差的两种增强应用。全局均值和方差是在整幅图像上计算的，这对于灰度和对比度的总体调整是有用的。这些参数的一种更有效的应用是局部增强，在局部增强中，局部均值和方差是根据图像中每个像素的邻域内的图像特征进行改变的基础。

令 (x, y) 表示给定图像中任意像素的坐标，S_{xy} 表示规定大小的以 (x, y) 为中心的邻域(子图像)。该邻域中像素的均值由下式给出：

$$m_{S_{xy}} = \sum_{i=0}^{L-1} r_i p_{S_{xy}}(r_i) \tag{3.3-22}$$

式中，$p_{S_{xy}}$ 是区域 S_{xy} 中像素的直方图。该直方图有 L 个分量，对应于输入图像中 L 个可能的灰度值。然而，许多分量是 0，具体取决于 S_{xy} 的大小。例如，如果邻域大小为 3×3 且 $L = 256$，那么该邻域的直方图的 256 个分量中仅 1 和 9 之间的分量非零。这些非零值对应 S_{xy} 中的不同灰度数(在 3×3 区域中可能的不同灰度的最大数是 9，最小数是 1)。

类似地，邻域中像素的方差由下式给出：

$$\sigma_{S_{xy}}^2 = \sum_{i=0}^{L-1} (r_i - m_{S_{xy}})^2 p_{S_{xy}}(r_i) \tag{3.3-23}$$

163~164

和前面一样，局部均值是邻域 S_{xy} 中平均灰度的度量，局部方差(或标准差)是邻域中灰度对比度的度

量。对邻域，我们可写出类似于式(3.3-20)和式(3.3-21)的表达式，即简单地对邻域中的像素值求和，并除以邻域中的像素数。

正如下面的例子说明的那样，使用局部均值和方差进行图像处理的一个重要方面是其灵活性，它们提供了简单而强有力的基于统计度量的增强技术，而统计度量与图像的外观有紧密的、可预测的关系。

例3.12　使用直方图统计的局部增强。

图 3.27(a)显示了一根绕在支架上的钨丝的 SEM(扫描电子显微镜)图像。图像中央的钨丝及其支架很清楚并很容易分析。在图像的右侧即图像的暗侧，有另一根钨丝的结构，但几乎看不到，其大小和特征当然也不容易辨认。通过对比度操作进行局部增强，是解决这种图像中包含部分隐含特征问题的理想方法。

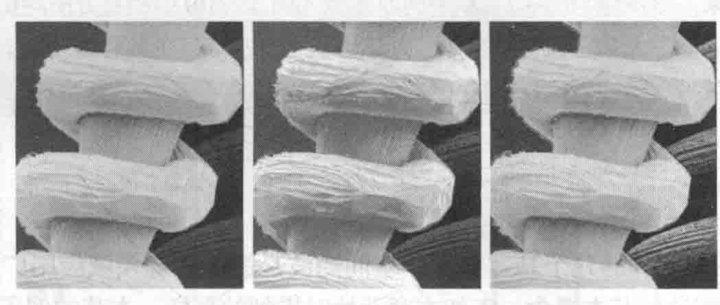


图 3.27　(a)放大了约 130 倍的钨丝的 SME 图像；(b)全局直方图均衡的结果；(c)使用局部直方图统计增强后的图像(原图像由俄勒冈大学地质科学系的 Michael Shaffer 先生提供)

在这种特殊情况中，问题是增强暗色区域，但同时尽可能使明亮区域不变，因为明亮区域并不需要增强。我们可以使用本节介绍的概念来阐明一种增强方法，这种方法能分辨暗区域与亮区域的不同，同时只增强暗区域。判断一个区域在点(x, y)是暗还是亮的方法是，把局部平均灰度$m_{S_{xy}}$与表示为m_G并称之为全局均值的平均图像灰度进行比较。这种量可使用整个图像并用式(3.3-18)或式(3.3-20)得到。这样，我们就有了增强方案的基础：如果$m_{S_{xy}} \leqslant k_0 m_G$，其中$k_0$是一个值小于 1.0 的正常数，那么我们将把点$(x, y)$处的像素考虑为处理的候选点。

因为我们感兴趣的是增强低对比度的区域，所以还需要一种度量方法来确定一个区域的对比度是否可作为增强的候选点。因此，若$\sigma_{S_{xy}} \leqslant k_2 \sigma_G$，则认为点$(x, y)$处的像素是增强的候选点，其中$\sigma_G$是由式(3.3-19)或式(3.3-21)得到的全局标准差，k_2为正常数。若我们的兴趣是增强亮区域，则该常数大于 1.0，对于暗区，则小于 1. 0。

最后，我们需要限制能够接受的最低对比度值，否则该过程会试图增强标准差为零的恒定区域。因此，我们通过要求$k_1 \sigma_G \leqslant \sigma_{S_{xy}}$，$k_1 < k_2$，对局部标准差设置一个较低的限制值。满足局部增强所有条件的一个位于点(x, y)处的像素，可简单地通过将像素值乘以一个指定常数E来处理，以便相对于图像的其他部分增大(或减小)其灰度值。不满足增强条件的像素则保持不变。

以上增强方法可总结如下：令$f(x, y)$表示在图像任意坐标(x, y)处的像素值，而令$g(x, y)$表示这些坐标处相应的增强的像素值，则对于$x = 0, 1, 2, \cdots, M-1, y = 0, 1, 2, \cdots, N-1$，有

$$g(x, y) = \begin{cases} E \cdot f(x, y), & m_{S_{xy}} \leq k_0 m_G \text{ 且 } k_1 \sigma_G \leq \sigma_{S_{xy}} \leq k_2 \sigma_G \\ f(x, y), & \text{其他} \end{cases} \tag{3.3-24}$$

式中，如前所述，E, k_0, k_1和k_2是规定的参数；m_G是输入图像的全局均值；σ_G是输入图像的标准差。参数$m_{S_{xy}}$和$\sigma_{S_{xy}}$分别是局部均值和标准差。按惯例，M和N是图像的行数和列数。

164 ~ 166

通常，式(3.3-24)中的选择参数需要做一些试验，以便熟悉一幅给定的图像或一类图像。在这种情况下，常选择下列数值：$E=4.0, k_0=0.4, k_1=0.02$ 和 $k_2=0.4$。之所以对 E 选择相对较低的值 4.0，原因在于当 E 与将被增强区域(暗色区域)中的灰度相乘后，结果仍会趋于灰度级的暗端，从而保持图像的视觉平衡。选择的 k_0 值比全局均值的一半略小，因为我们观察一幅图像时，我们了解到需要增强的区域确实比全局平均值的一半还要暗。类似的分析可指导 k_1 和 k_2 值的选择。这些常数的选择通常并不困难，但这些明确的选择必须以对具体增强问题的逻辑分析为指导。最后，为保留细节和尽量减少计算负担，局部区域 S_{xy} 的大小应尽可能小。因此，我们选择大小为 3×3 的小区域。

作为比较的基础，我们使用全局直方图均衡增强一幅图像。图3.27(b)显示了这一结果。暗区域已被改进，但分辨细节还是很困难，亮区域也有变化，而这是我们不希望的结果。图3.27(c)显示了使用上面描述的局部统计方法处理的结果。该图像与图 3.27(a)中的原图像或图 3.27(b)中经直方图均衡后的结果相比，我们注意到，图3.27(c)的右边显示出了明显的细节。观察该例子，暗灯丝中的脊线现在非常清楚。值得注意的是，左侧的亮灰度区域被完整地保留，而这正是我们的初始目的之一。

3.4 空间滤波基础

本节介绍空间滤波图像处理的几个基本概念。空间滤波是图像处理领域应用广泛的主要工具之一。因此，我们应充分理解这些基本概念。正如本章开始时提到的那样，本节的例子多数涉及使用空间滤波来增强图像。空间滤波的其他应用将在后面几章中讨论。

滤波一词借用于频率域处理，频率域处理是下一章的主题，在下一章中，“滤波”是指接受(通过)或拒绝一定的频率成分。例如，通过低频的滤波器称为低通滤波器。低通滤波器的最终效果是模糊(平滑)一幅图像。我们可以用空间滤波器(也称为空间掩模、核、模板和窗口)直接作用于图像本身而完成类似的平滑。事实上，如第 4 章所示，线性空间滤波与频率域滤波之间存在一一对应关系。然而，空间滤波可提供相当多的功能，就像我们稍后将要看到的那样，它们还可用于非线性滤波，而这在频率域中是做不到的。

关于线性的讨论，请参阅 2.6.2 节。

3.4.1 空间滤波机理

在图 3.1 中，我们简单解释过，空间滤波器由(1)一个邻域(通常是一个较小的矩形)，(2)对该邻域所包围图像像素执行的预定义操作组成。滤波产生一个新像素，新像素的坐标等于邻域中心的坐标，像素的值是滤波操作的结果[①]。滤波器的中心访问输入图像中的每个像素后，就生成了处理(滤波)后的图像。如果在图像像素上执行的是线性操作，则该滤波器称为线性空间滤波器。否则，滤波器就称为非线性空间滤波器。首先，我们重点关注线性滤波器，然后说明某些简单的非线性滤波器。5.3 节将包含更多关于非线性滤波器及其应用的内容。

图 3.28 说明了使用 3×3 邻域的线性空间滤波的机理。在图像中的任意一点 (x,y)，滤波器的响应 $g(x,y)$ 是滤波器系数与由该滤波器所包围的图像像素的乘积之和：

$$g(x,y)=w(-1,-1)f(x-1,y-1)+w(-1,0)f(x-1,y)+\cdots+w(0,0)f(x,y)+\cdots+w(1,1)f(x+1,y+1)$$

很明显，滤波器的中心系数 $w(0,\ 0)$ 对准位置 (x,y) 的像素。对于大小为 $m\times n$ 的模板，我们假设 $m=2a+1$ 且 $n=2b+1$，其中 a, b 为正整数。这意味着在后续的讨论中，我们关注的是奇数尺寸的滤

166 ~ 167

① 滤波后的像素值通常会赋给新创建图像中的对应位置，以容纳滤波的结果。滤波后的像素代替原图像中对应位置的值的情形很少见，因为这会在改变图像内容的同时执行滤波操作。

波器，其最小尺寸为3×3。一般来说，使用大小为$m\times n$的滤波器对大小为$M\times N$的图像进行线性空间滤波，可由下式表示：

$$g(x,y)=\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)f(x+s,y+t)$$

式中，x和y是可变的，以便w中的每个像素可访问f中的每个像素。

当然，也可以使用偶数尺寸的滤波器，或使用混合有偶数尺寸和奇数尺寸的滤波器。但是，使用奇数尺寸的滤波器可简化索引，并更为直观，因为滤波器的中心落在整数值上。

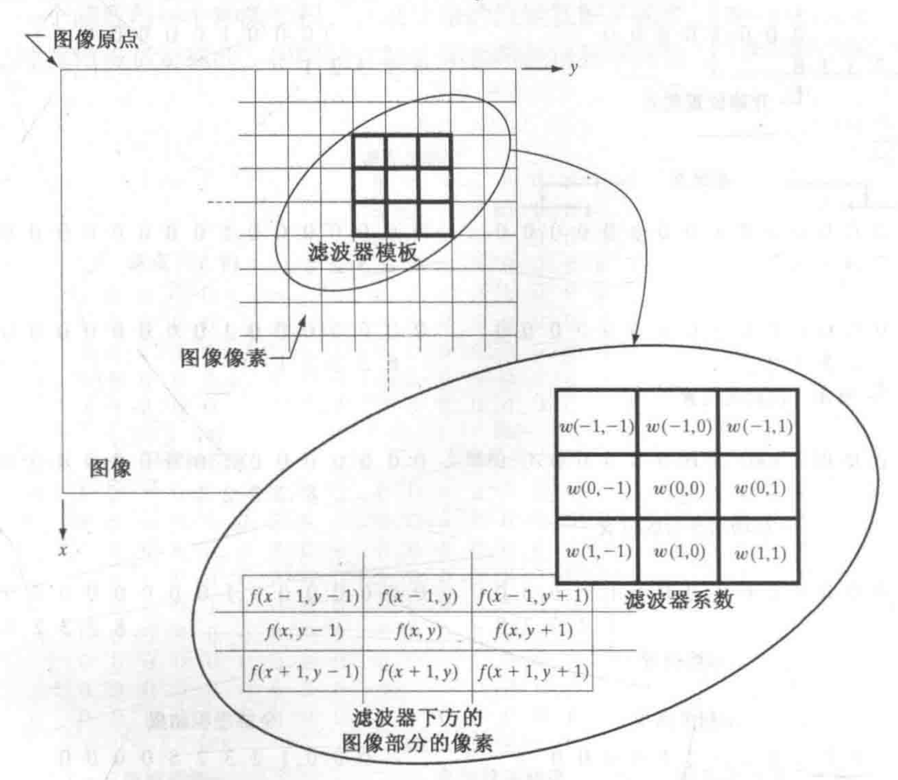


图 3.28　使用大小为3×3的滤波器模板的线性空间滤波的机理。表示滤波器模板系数的坐标所选择的形式简化了线性滤波的表达式

3.4.2　空间相关与卷积

在执行线性空间滤波时，必须清楚地理解两个相近的概念。一个是相关，另一个是卷积。如前节描述的那样，相关是滤波器模板移过图像并计算每个位置乘积之和的处理。卷积的机理相似，但滤波器首先要旋转180°。解释这两个概念的区别的最好方法是举一个例子。我们从一个一维示例开始。

图3.29(a)显示了一个一维函数f和一个滤波器w，图3.29(b)显示了执行相关的起始位置。我们注意到的第一件事情是存在未覆盖的部分函数。该问题的解决办法是在f的任意一侧补上足够的0，以便使w中的每个像素都可访问到f中的每个像素。如果滤波器的尺寸是m，那么需要在f的一侧补$m-1$个0。图3.29(c)显示了适当填充过的函数。相关的第一个值是如图3.29(c)所示的初始位置的f和w的乘积之和(乘积之和为0)。这相当于位移$x=0$。为了得到相关的第二个值，我们把w向右移动一个像素的位置(位移$x=1$)，并计算乘积之和。结果还是0。事实上，当$x=3$时才第一次出现非零的结果，在这种情况下，w中的8覆盖f中的1，相关的结果是8。按照这种方法进行处理，我们可以得到图3.29(g)中的全部相关结果。注意，x取了12个值(即$x=0,1,2,\cdots,11$)使w滑过f，以便w中的每个像素访问f中的每个像素。

零填充并不是唯一的选择。例如，我们可在f的两侧复制第一个元素和最后一个元素$m-1$次，或镜像第一个元素和最后一个元素$m-1$次，并为填充使用镜像后的值。

167~168

通常，我们喜欢使用大小与 f 相同的相关阵列，在这种情况下，我们将全部相关结果裁剪到原函数大小，如图3.29(h)所示。



图 3.29　具有离散单位冲激的滤波器的一维相关与卷积的说明

从前一段的讨论可知，有两个要点需要注意。第一，相关是滤波器位移的函数。换句话说，相关的第一个值对应于滤波器的零位移，第二个值对应于一个单元位移，等等。第二，滤波器 w 与包含有全部 0 和单个 1 的函数相关，得到的结果是 w 的一个副本，但旋转了180°。我们将包含单个 1 而其余都是 0 的函数称为离散单位冲激。因此，我们得出结论：一个函数与离散单位冲激相关，会在该冲激位置产生该函数的一个翻转版本。

卷积的概念是线性系统理论的基石。正第 4 章介绍的那样，卷积的基本特性是某个函数与某个单位冲激卷积，得到函数在该冲激处的一个副本。在前一段中我们看到，相关操作也得到函数的一个副本，但该副本旋转了180°。因此，如果我们预先旋转滤波器，并执行相同的滑动乘积求和操作，就应该能得到希望的结果。如图3.29右边一列所示的那样，情况的确如此。由此，我们看到，为了执行卷积，我们需要做的是把一个函数旋转180°，然后执行相关中的相同操作。如结果所示，我们旋转两个函数的做法没有区别。

> 注意，旋转180°等同于水平翻转该函数。

168 ~ 169

如图 3.30 所示，前面的概念很容易扩展到图像。对于大小为 $m\times n$ 的滤波器，我们在图像的顶部和

底部至少填充 $m-1$ 行 0，在左侧和右侧填充 $n-1$ 列 0。在这种情况下，m 和 n 等于 3，因此，如图3.30(b)所示，我们用两行 0 填充图像的顶部和底部，用两列 0 填充图像的左侧和右侧。图3.30(c)显示了执行相关操作的滤波器模板的初始位置，图3.30(d)显示了所有相关操作的结果。图3.30(e)显示了裁剪后的相应结果。再一次注意到，结果旋转了180°。对于卷积，与前面一样，我们预先旋转模板，然后使用刚才描述的方法对乘积做滑动求和操作。图3.30(f)到(h)显示了结果。我们再次看到，一个函数与一个冲激卷积，在该冲激的位置复制了这个函数。如果滤波器模板是对称的，很明显，相关和卷积将得到相同的结果。

> 在二维情况下，旋转 180° 等同于沿一个坐标轴翻转模板，然后沿另一个坐标轴再次翻转模板。

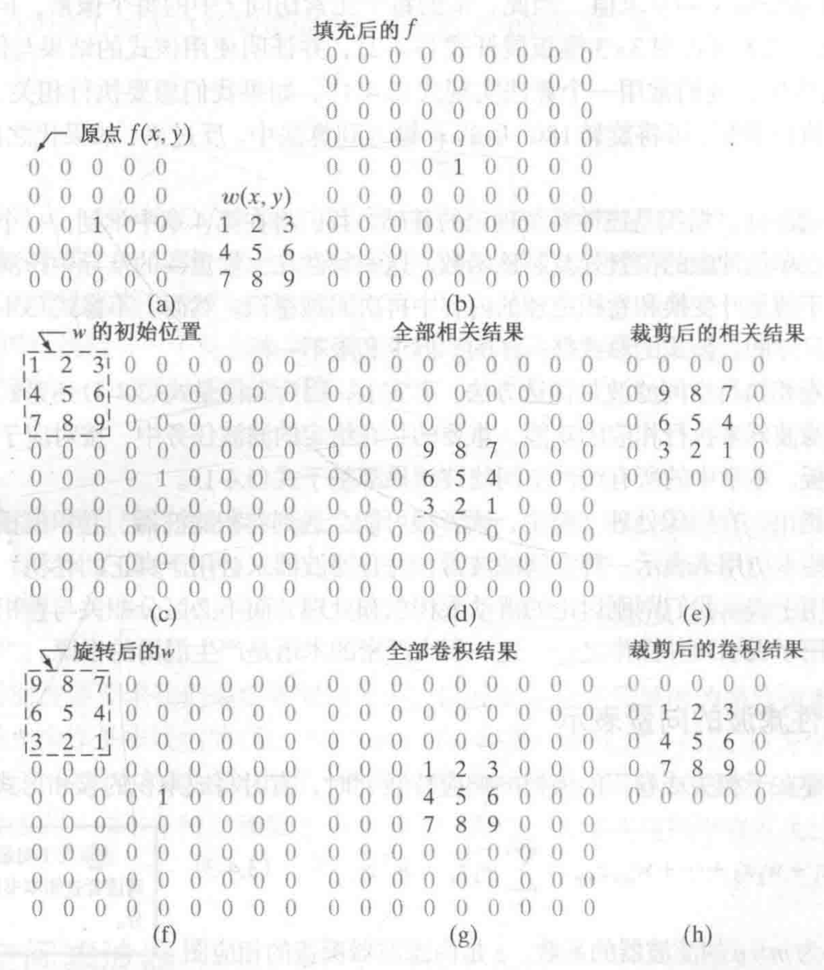


图 3.30　二维滤波器与二维离散单位冲激的相关(中间一行)和卷积(最后一行)。为便于视觉分析，0 显示为灰色

替代包含单个 1，如果图 3.30 中的图像 f 包含一个与 w 完全相等的区域，当 w 位于 f 的区域的中心时，相关函数(归一化后)的值将是最大的。这样，正像我们将在第 12 章看到的那样，相关还可以用于寻找图像中的匹配。

下面以公式形式总结一下前面的讨论。一个大小为 $m \times n$ 的滤波器 $w(x, y)$ 与一幅图像 $f(x, y)$ 做相关操作，可表示为 $w(x, y) \star f(x, y)$，它由上一节末尾列出的公式给出，为方便起见我们重写如下：

$$w(x, y) \star f(x, y)=\sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s, t) f(x+s, y+t) \tag{3.4-1}$$

170 ~ 171

这一等式对所有位移变量 x 和 y 求值，以便 w 的所有元素访问 f 的每个像素，其中我们假设 f 已被

适当地填充。正如早些时候解释的那样，$a=(m-1)/2, b=(n-1)/2$，同时，为表示方便，我们假设 m 和 n 是奇整数。

类似地，$w(x,y)$ 和 $f(x,y)$ 的卷积表示为 $w(x,y)\bigstar f(x,y)$[①]，它由下式给出：

$$w(x,y)\bigstar f(x,y)=\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)f(x-s,y-t) \tag{3.4-2}$$

> 通常，在含义明确的情况下，我们用函数 $g(x,y)$ 而不用 $w(x,y)\star f(x,y)$ 或 $w(x,y)\bigstar f(x,y)$ 来表示相关或卷积的结果。例如，参见前一节末尾的等式和式(3.5-1)。

其中等式右侧的减号表示翻转 f(即旋转180°)。为简化符号表示，我们遵循惯例，翻转和移位 w 而非 f。结果是一样的。与相关一样，该式也对所有位移变量 x 和 y 求值，因此，w 的每个元素访问 f 中的每个像素，同样我们也假设 f 已被适当地填充。读者可以对 3×3 模板展开式(3.4-2)，并证明使用该式的结果与图 3.30 中例子的结果相同。在实践中，我们常用一个算法实现式(3.4-1)。如果我们想要执行相关，可将 w 输入到算法中；如果要执行卷积，可将旋转180°后的 w 输入到算法中。反过来，如果代之以执行式(3.4-2)也可行。

如之前提及的那样，卷积是线性系统理论的基础。我们将在第 4 章中学到，一个函数与单位冲激的卷积，相当于在单位冲激的位置处复制该函数，这一特性在大量重要的推导中扮演核心的角色。我们将在第 4 章关于傅里叶变换和卷积定理的内容中再次回顾卷积。然而，不像式(3.4-2)，我们将要处理的卷积是相同尺寸的。公式的形式是一样的，但求和限不一样。

使用相关或卷积执行空间滤波是优选方法。事实上，因为无论是式(3.4-1)还是式(3.4-2)，都可以通过简单地旋转滤波器来执行相应的功能，重要的是在给定的滤波任务中，按对应于期望操作的方式来指定滤波器模板。本章中的所有线性空间滤波结果都基于式(3.4-1)。

最后，我们指出，在图像处理文献中，读者很可能会遇到卷积滤波器、卷积模板或卷积核这些术语。按惯例，这些术语用来表示一种空间滤波器，并且滤波器未必用于真正的卷积。类似地，模板与图像的卷积通常用于表示我们刚刚讨论的滑动乘积求和处理，而不必区分相关与卷积间的差别。更合适的是，它通常用于表示两种操作之一。这一不太严密的术语是产生混淆的根源。

3.4.3 线性滤波的向量表示

当我们的兴趣在于相关或卷积的模板的响应特性 R 时，有时写成乘积的求和形式是方便的：

$$R=w_1z_1+w_2z_2+\cdots+w_{mn}z_{mn}=\sum_{k=1}^{mn}w_kz_k=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{z} \tag{3.4-3}$$

> 关于向量和矩阵的基本知识，请读者查阅本书网站上的 Tutorials 部分。

式中 w 项是大小为 $m\times n$ 的滤波器的系数，z 是由滤波器覆盖的相应图像的灰度值。如果兴趣是使用式(3.4-3)来做相关，则可用给定的模板。为使用相同的公式进行卷积操作，如上节描述的那样，我们可以简单地把模板旋转180°。它意味着式(3.4-3)对特定的坐标对 (x,y) 是成立的。在下一节中，我们将了解为什么这一表示法对于描述给定的线性滤波器的特性是方便的。

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

图 3.31 普通 3×3 滤波器模板的另一种表示

作为一个例子，图3.31 显示了一个带有标号的普通 3×3 模板。在这种情况下，式(3.4-3)变为

171 ~ 173

① 因为卷积满足交换律，故有 $w(x,y)\bigstar f(x,y)=f(x,y)\bigstar w(x,y)$，相关则不满足交换律，这一点可通过颠倒图3.29(a)中函数的顺序看出。

$$R = w_1 z_1 + w_2 z_2 + \cdots + w_9 z_9 = \sum_{k=1}^{9} w_k z_k = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{z} \tag{3.4-4}$$

式中，$\boldsymbol{w}$ 是由模板的系数形成的 9 维向量，$\boldsymbol{z}$ 是由模板包含的图像灰度形成的 9 维向量。

3.4.4 空间滤波器模板的产生

生成一个大小为 $m \times n$ 的线性空间滤波器要求指定 mn 个模板系数，这些系数是根据该滤波器支持什么样的操作来选择的，记住，线性滤波的作用实现乘积求和操作。例如，假设我们想要将图像中的这些像素替换为以这些像素为中心的3×3 邻域的平均灰度。图像中任意位置 (x, y) 的灰度平均值是以 (x, y) 为中心的3×3 邻域中的 9 个灰度值之和除以 9。令 $z_i, i = 1, 2, \cdots, 9$ 表示这些灰度，则平均灰度为

$$R = \frac{1}{9} \sum_{i=1}^{9} z_i$$

但该式与系数值为 $w_i = 1/9$ 的式(3.4-4)相同。换句话说，使用系数为 1/9 的3×3 模板进行线性滤波操作可实现所希望的平均。正如下一节中将要讨论的那样，这一操作将导致图像平滑。在下面几节中，我们讨论基于这种方法的其他几个滤波器模板。

在某些应用中，我们有一个具有两个变量的连续函数，目的是基于该函数得到一个空间滤波模板。例如，有两个变量的一个高斯函数有如下基本形式：

$$h(x, y) = \mathrm{e}^{-\frac{x^2 + y^2}{2\sigma^2}}$$

式中，σ 是标准差，并且我们通常假设坐标 x 和 y 是整数。譬如，为了从该函数产生一个大小为3×3 的滤波器模板，我们关于其中心进行取样。因此有 $w_1 = h(-1, -1),\ w_2 = h(-1, 0), \cdots,\ w_9 = h(1, 1)$。使用类似的方式可产生一个 $m \times n$ 滤波器模板。回顾可知，二维高斯函数具有钟形形状，并且其标准差控制钟形的“紧度”。

产生非线性滤波器要求我们确定邻域的大小，以及对包含在邻域内的图像像素执行的操作。例如，回忆可知最大操作是非线性的(见 2.6.2 节)，以一幅图像中任意一点 (x, y) 为中心的一个 5×5 最大滤波器得到 25 个像素中的最大灰度值，并将该值赋给处理后图像中位置为 (x, y) 的像素。非线性滤波器功能非常强大，如本章稍后和第 5 章中显示的那样，在某些应用中它可执行超出线性滤波器能力的功能。

3.5 平滑空间滤波器

平滑滤波器用于模糊处理和降低噪声。模糊处理经常用于预处理任务中，例如在(大)目标提取之前去除图像中的一些琐碎细节，以及连接直线或曲线的缝隙。通过线性滤波和非线性滤波模糊处理，可以降低噪声。

3.5.1 平滑线性滤波器

平滑线性空间滤波器的输出(响应)是包含在滤波器模板邻域内的像素的简单平均值。这些滤波器有时也称为均值滤波器。如前节提到的那样，也可以把它们归入低通滤波器。

平滑滤波器的基本概念非常直观。它使用滤波器模板确定的邻域内像素的平均灰度值来代替图像中每个像素的值，这种处理的结果降低了图像灰度的“尖锐”变化。由于典型的随机噪声由灰度级

173 ~ 174

的急剧变化组成，因此，常见的平滑处理应用就是降低噪声。然而，由于图像边缘(几乎总是一幅图像希望有的特性)也是由图像灰度尖锐变化带来的特性，所以均值滤波处理还是存在着不希望有的边缘模糊的负面效应。另外，这类处理的其他应用包括如2.4.3节讨论过的由于灰度级数量不足而引起的伪轮廓效应的平滑处理。均值滤波器的主要应用是去除图像中的不相关细节，其中“不相关”是指与滤波器模板尺寸相比较小的像素区域。这将在本节稍后的应用中加以说明。

图3.32显示了两个 3×3 平滑滤波器。第一个滤波器产生模板下方的标准像素平均值。把模板系数代入式(3.4-4)即可清楚地看出这一点：

$$R=\frac{1}{9}\sum_{i=1}^{9}z_i$$

R 是由模板定义的 3×3 邻域内像素灰度的平均值。注意，代替上式中的1/9，滤波器的系数全为“1”。这里的概念是系数取1值时计算更有效。在滤波处理之后，整个图像除以9。一个 $m\times n$ 模板应有等于 $1/mn$ 的归一化常数。所有系数都相等的空间均值滤波器有时称为盒状滤波器。

图3.32所示的第二个模板更为重要。该模板产生所谓的加权平均，使用这一术语是指用不同的系数乘以像素，即一些像素的重要性(权重)要比另一些像素的重要性大。在图3.32(b)所示的模板中，处于该模板中心位置的像素所乘的值比其他任何像素所乘的值都要大，因此，在均值计算中为该像素提供更大的重要性。其他像素如同是模板中心距离的函数那样赋以成反比的权重。由于对角项离中心比离正交方向相邻的像素(参数为 $\sqrt{2}$)更远，因此其权重比与中心直接相邻的像素更小。赋予中心点最高权重，然后随着距中心点距离的增加而减小系数值的加权策略的目的，是在平滑处理中试图降低模糊。我们也可选择其他权重来达到相同的目的。然而，图3.32(b)所示模板中所有系数的和等于16，对于计算机计算来说是一个有吸引力的特性，因为它是2的整数次幂。在实践中，由于这些模板在一幅图像中任何一个位置所跨过的区域很小，通常很难看出使用图3.32中的各种模板或类似方式进行平滑处理后的图像之间的区别。

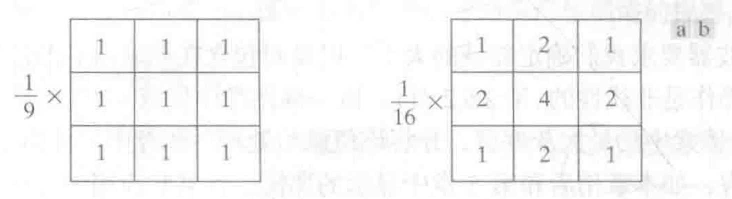


图3.32 两个 3×3 平滑(均值)滤波器模板。每个模板前面的乘数等于1除以所有系数之和，这是计算平均值所要求的

参见式(3.4-1)，一幅 $M\times N$ 图像经过一个 $m\times n$ (m 和 n 是奇数)的加权均值滤波器滤波的过程可由下式给出：

$$g(x,y)=\frac{\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)f(x+s,y+t)}{\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)} \tag{3.5-1}$$

上式中的参数见式(3.4-1)中的定义。如前所述，它可以这样理解，即一幅完全滤波的图像是通过对 $x=0,1,2,\cdots,M-1$ 和 $y=0,1,2,\cdots,N-1$ 执行式(3.5-1)得到的。式(3.5-1)中的分母部分简单地表示为模板的各系数之和，它是一个仅需计算一次的常数。

174~175

例 3.13　使用各种尺寸的模板平滑图像。

与滤波器模板尺寸有关的图像平滑效果如图 3.33 所示，图中显示了一幅原图像以及分别用尺寸为 $m = 3$, 5, 9, 15 和 35 像素的方形均值滤波器得到的相应平滑结果。这些结果的主要特点如下：当 $m = 3$ 时，可以观察到在整幅图像中有轻微的模糊，但正如所希望的那样，当图像细节与滤波器模板近似相同时，图像中一些细节受到的影响比较大。例如，图像中的 3×3 与 5×5 黑方块、较小的字母 “a” 和细颗粒噪声，与图像的其他部分相比，要更模糊一些。注意，噪声明显降低了，字母的锯齿状边缘也令人满意地平滑了。

$m = 5$ 时的结果基本类似，但模糊程度稍微有所增加。在 $m = 9$ 时，可以明显地看出图像更加模糊，而且有 20%的黑色圆圈几乎不能像在前三幅图像中那样与背景明显地区分开来，这表明当目标的灰度与其相邻像素的灰度接近时，会导致模糊目标的混合效应。$m = 15$ 和 35 时，就图像中目标物的尺寸来说，已属极端情况。这种极端类型的模糊处理通常用于去除图像中的一些小物体。例如，在图 3.35(f) 中，三个小方框、两个小圆圈及大部分噪声矩形区域已融入背景。注意，在这幅图中还有明显的黑边界。这是我们用 0(黑色) 填充原图像的边界，经过滤波后再去除填充区域的结果。某些黑色混入了所有滤波后的图像，但对于使用较大滤波器平滑的图像，这就会成为问题。

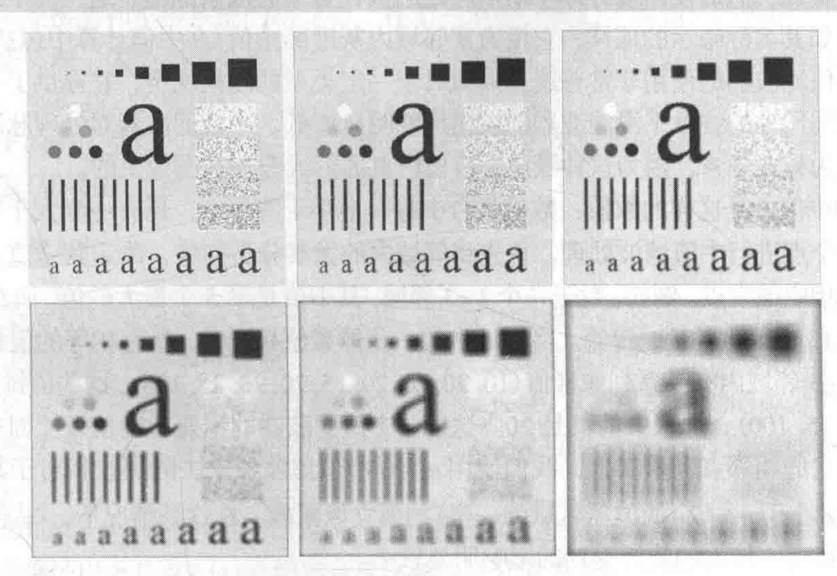
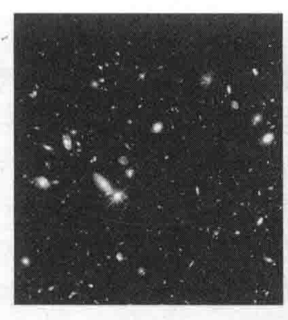
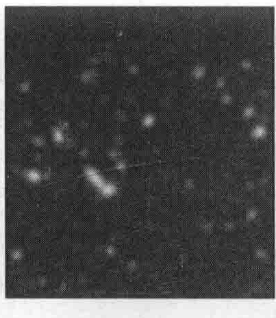
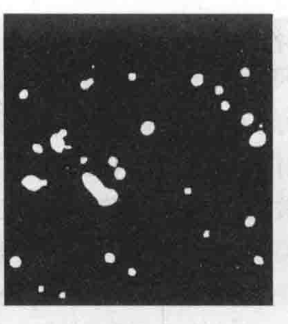


图 3.33　(a) 大小为 500×500 像素的原图像；(b)~(f) 用大小分别为 $m = 3$, 5, 9, 15 和 35 的方形均值滤波模板平滑处理的结果。顶部的黑色方块大小分别为 3, 5, 9, 15, 25, 35, 45 和 55 个像素，它们的边界相隔 25 个像素。底部字母的大小在 10 磅到 24 磅之间，增量为 2 磅；顶部的最大字母为 60 磅。垂直线段条宽为 5 个像素，高为 100 个像素，线条间距为 20 个像素。圆的直径为 25 个像素，它们的边界相隔 15 个像素，灰度级为 0%到 100%的黑色范围内，以 20%增加。图像背景为黑色的 10%。噪声矩形区域大小为 50×120 像素

如早些时候提到的那样，空间均值处理的一个重要应用是，模糊图像以便得到感兴趣物体的粗略描述，因此那些较小物体的灰度就会与背景混合在一起，较大物体则变得像 “斑点” 而易于检测。模板的大小由那些即将融入背景中的物体尺寸决定。作为实例，考虑图3.34(a)，它是绕地轨道上的哈勃望远镜拍摄的一幅图像。图3.34(b) 显示了应用 15×15 均值滤波器模板对该图像处理后的结果。可以看到，图像中的一些部分要么已融入背景，要么亮度明显降低。像这样利用阈值处理并基于物体亮度来消除某些物体的操作是很典型的。使用等于模糊图像最高亮度 25%的阈值，并用图3.2(b) 的阈值函数处理的结果示于图3.34(c) 中。将该结果与原图像比较，我们看到，考虑图像中最大、最亮物体的表达方式是合理的。

176 ~ 177

a b c

图 3.34　(a)来自哈勃太空望远镜的大小为 528×485 像素的图像；(b)由 15×15 均值模板滤波后的图像；(c)对图(b)进行阈值处理后的结果(原图像由 NASA 提供)

3.5.2　统计排序(非线性)滤波器

统计排序滤波器是一种非线性空间滤波器，这种滤波器的响应以滤波器包围的图像区域中所包含的像素排序(排队)为基础，然后使用统计排序结果决定的值代替中心像素的值。这一类中最知名的滤波器是中值滤波器，如其名称暗示的那样，它用像素邻域内灰度的中值(在中值计算中包括原像素值)代替该像素的值。中值滤波器的使用非常普遍，因为对于一定类型的随机噪声，它提供了一种优秀的去噪能力，而且比相同尺寸的线性平滑滤波器的模糊程度明显要低。中值滤波器对处理脉冲噪声非常有效，这种噪声也称为*椒盐噪声*，因为这种噪声是以黑白点的形式叠加在图像上的。

一个数值集合的中值 ξ 是这样的数值：数值集合中有一半小于等于 ξ，还有一半大于等于 ξ。为了对一幅图像上的某个点进行中值滤波处理，首先将邻域内的像素分类排序，确定其中值，并将中值赋予滤波后图像中的相应像素点。例如，对于一个 3×3 邻域，其中值是第 5 个最大的值，而在一个 5×5 邻域中，中值就是第 13 个最大的值，等等。当邻域中的一些像素值相同时，所有相等的值都可以作为中值。假如，在一个 3×3 邻域内有一系列像素值(10, 20, 20, 20, 15, 20, 20, 25, 100)，这些值排序后为(10, 15, 20, 20, 20, 20, 20, 25, 100)，那么中值就是 20。这样，中值滤波器的主要功能就是使拥有不同灰度的点看起来更接近于它的相邻点。事实上，我们使用 $m \times m$ 中值滤波器来去除那些相对于其邻域像素更亮或更暗，且其区域小于 $m^2/2$ (滤波器区域的一半)的孤立像素族。在这种情况下，“去除”的意思是强制为邻域的中值灰度。较大的族所受到的影响明显较小。

在图像处理中，尽管中值滤波器是使用得最为广泛的统计排序滤波器，但并不意味着它是唯一的。中值象征一系列像素值排序后的第 50%个值，但根据基本统计学可知，排序也适用于其他不同的情况。例如，可以取第 100%个值，即所谓的最大值滤波器，这种滤波器在搜寻一幅图像中的最亮点时非常有用。3×3 最大值滤波器的响应可以由式 $R = \max\{z_k \mid k = 1, 2, \cdots, 9\}$ 给出。相反，取第 0%个值的滤波器就是所谓的最小值滤波器，它可用于相反的目的。中值、最大值、最小值和其他非线性滤波器将在 5.3 节中详细介绍。

关于百分比的内容，请参阅 10.3.5 节。

例 3.14　利用中值滤波降噪。

图 3.35(a)显示了一幅被椒盐噪声污染的电路板的 X 射线图像。为了说明这种情况下的中值滤波器处理效果比均值滤波器更好，我们在图 3.35(b)中显示了用 3×3 邻域均值模板处理噪声图像的结果，而在图 3.35(c)中显示了用 3×3 中值滤波器处理噪声图像的结果。均值滤波模糊了图像，并且噪声去除性能也很差。在这种情况下，中值滤波的性能要远优于均值滤波。通常，中值滤波要比均值滤波更适合去除椒盐噪声。

178 ~ 179

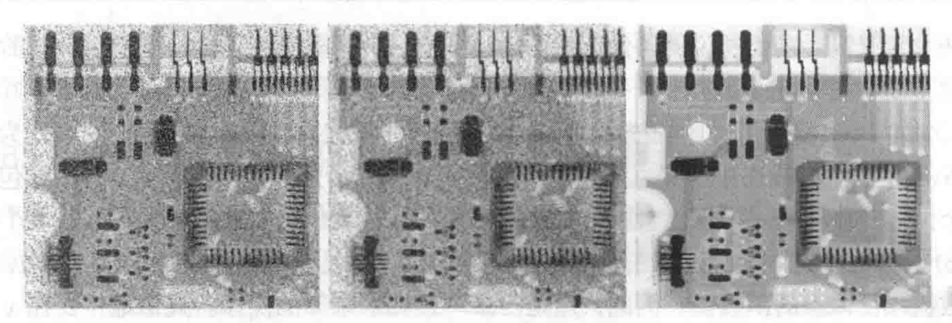


图 3.35　(a)被椒盐噪声污染的电路板的 X 射线图像；(b)用 3×3 均值模板降噪后的图像；(c)用 3×3 中值滤波器降噪后的图像(原图像由 Lixi 公司的 Joseph E. Pascente 先生提供)

3.6　锐化空间滤波器

锐化处理的主要目的是突出灰度的过渡部分。图像锐化的用途多种多样，应用范围从电子印刷和医学成像到工业检测和军事系统的制导等。在上一节中，我们了解到，图像模糊可通过在空间域用像素邻域平均法实现。因为均值处理与积分类似，在逻辑上，我们可以得出锐化处理可由空间微分来实现这一结论。事实上，的确如此，且本节将讨论由数字微分来定义和实现锐化算子的各种方法。基本上，微分算子的响应强度与图像在用算子操作的这一点的突变程度成正比，这样，图像微分会增强边缘和其他突变(如噪声)，削弱灰度变化缓慢的区域。

3.6.1　基础

以下两节分别详细讨论基于一阶微分和二阶微分的锐化滤波器。在讨论具体滤波器之前，先回顾数字意义上微分的某些基本性质。为简化说明，主要集中讨论一阶微分的性质。我们最感兴趣的是恒定灰度区域中，突变的起点与终点(台阶和斜坡突变)及灰度斜坡处的微分性质。如第 10 章介绍的那样，这些类型的突变可用来对图像中的噪声点、线与边缘建模。这些图像特征过渡处的微分性质也很重要。

数字函数的微分可用不同的术语定义，也存在定义这些差别的各种方法。然而，对于一阶微分的任何定义都必须保证以下几点：(1)在恒定灰度区域的微分值为零；(2)在灰度台阶或斜坡处的微分值非零；(3)沿斜坡的微分值非零。类似地，任何二阶微分的定义必须保证以下几点：(1)在恒定区域微分值为零；(2)在灰度台阶或斜坡的起点处微分值非零；(3)沿斜坡的微分值非零。因为我们处理的是数字量，其值是有限的，所以最大灰度级的变化也是有限的，并且变化发生的最短距离是在两相邻像素之间。

一维函数 $f(x)$ 的一阶微分的基本定义是差值

$$\frac{\partial f}{\partial x} = f(x+1) - f(x) \tag{3.6-1}$$

> 10.2.1 节将会回到式(3.6-1)，并给出由泰勒级数推导出它的方法。现在，我们只须把它当作一个定义来接受即可。

式中，为了与二维图像函数 $f(x,y)$ 的微分保持一致，我们使用了偏导数符号。对于二维函数，我们将沿两个空间轴来处理偏微分。当前讨论的空间微分应用并不影响我们试图采用的任何方法的本质。很明显，当函数中只有一个变量时，$\partial f/\partial x = \mathrm{d}f/\mathrm{d}x$；对于二阶微分，这同样也成立。

我们将二阶微分定义为如下差分：

$$\frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x) \tag{3.6-2}$$

179
~
180

很容易验证这两个定义满足前面所说的条件。为了解这一点，并考察数字函数一阶微分和二阶微分间的异同点，考虑图 3.36 中的例子。

图3.36(b)(图的中间)显示了一段扫描线。小方块中的数值是扫描线中的灰度值，它们作为黑点画在上方的图3.36(a)中。用虚线连接这些点是为了帮助我们看得更清楚。如图中所示，扫描线包含一个灰度斜坡、三个恒定灰度段和一个灰度台阶。圆圈指出了灰度变化的起点和终点。用前面两个定义计算出的图3.36(b)中扫描线的一阶微分和二阶微分画在图3.36(c)中。计算点 x 处的一阶微分时，我们用下一个点的函数值减去该点的函数值。因此，这是一个"预测未来"的操作。类似地，要在 x 点计算二阶微分，计算中要使用前一个点和下一个点。为避免前一个点和下一个点处于扫描线之外的情况，我们在图 3.36 中显示了从序列中第二个点到倒数第二个点的微分计算。

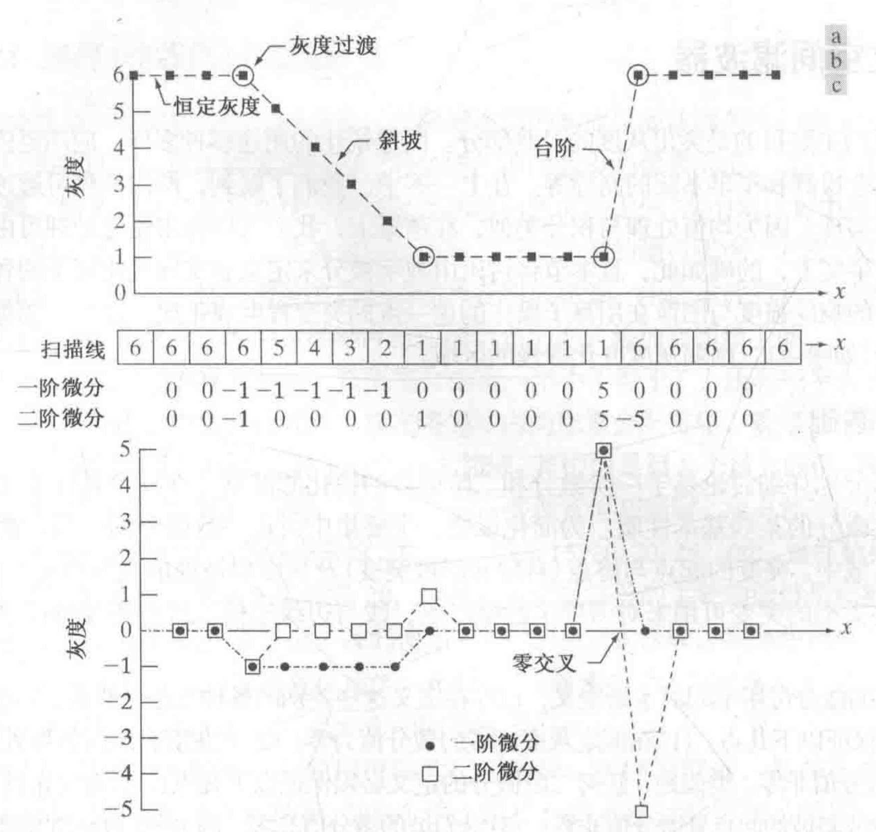


图 3.36 一幅图像中一段水平灰度剖面的一维数字函数的一阶微分和二阶微分的说明。在图(a)和图(c)中，为便于观看，已用虚线将数据点连接起来

让我们从左到右横贯剖面图，考虑一阶微分和二阶微分的性质。首先，我们遇到的是如图3.36(b)和(c)所示的恒定灰度区域，其一阶微分和二阶微分都是零，因此两者都满足条件(1)。接着，我们遇到紧随台阶的一个灰度斜坡，并注意到在斜坡起点和台阶处的一阶微分不为零，类似地，在斜坡和台阶的起点和终点的二阶微分也不为零；因此，两个微分特性都满足条件(2)。最后，我们看到，两个微分特性也都满足条件(3)，因为对于斜坡来说一阶微分不是零，二阶微分是零。请注意斜坡或台阶的起点和终点处二阶微分的符号变化。事实上，我们在图 3.36(c)中看到，在一个台阶的过渡部分，连接这两个值的线段在两个端点的中间与水平轴相交。如第 10 章所述，零交叉对于边缘定位是非常有用的。

181 ~ 182 数字图像中的边缘在灰度上通常类似于斜坡过渡，因此会导致图像的一阶微分产生较粗的边缘，因为沿斜坡的微分非零。另一方面，二阶微分产生由零分开的一个像素宽的双边缘。由此，

我们得出结论，二阶微分在增强细节方面要比一阶微分好得多，是一个适合锐化图像的理想特性。如本节稍后所述，二阶微分比一阶微分执行上要容易得多，因此我们重点介绍二阶微分。

3.6.2　使用二阶微分锐化图像——拉普拉斯算子

本节介绍二维函数的二阶微分实现及其在图像锐化处理中的应用。在第 10 章中，我们将再次回到微分问题上来，其时将广泛使用二阶微分来分割图像。这种方法基本上是先定义一个二阶微分的离散公式，然后根据公式构造一个滤波器模板。我们重点介绍一种各向同性滤波器，这种滤波器的响应与滤波器作用的图像的突变方向无关。也就是说，各向同性滤波器是旋转不变的，即将原图像旋转后进行滤波处理给出的结果，与先对图像滤波然后再旋转的结果相同。

可以证明(Rosenfeld and Kak[1982])，最简单的各向同性微分算子是拉普拉斯算子。一个二维图像函数 $f(x,y)$ 的拉普拉斯算子定义为

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{3.6-3}$$

因为任意阶微分都是线性操作，所以拉普拉斯变换也是一个线性算子。为了以离散形式描述这一公式，我们使用式(3.6-2)的定义，记住，我们必须支持第二个变量。在 x 方向上，有

$$\frac{\partial^2 f}{\partial x^2} = f(x+1,y) + f(x-1,y) - 2f(x,y) \tag{3.6-4}$$

类似地，在 y 方向上有

$$\frac{\partial^2 f}{\partial y^2} = f(x,y+1) + f(x,y-1) - 2f(x,y) \tag{3.6-5}$$

所以，满足这三个公式的两个变量的离散拉普拉斯算子是

$$\nabla^2 f(x,y) = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y) \tag{3.6-6}$$

这个公式可以用图3.37(a)的滤波模板来实现，该图给出了以 90°为增量进行旋转的各向同性结果。实现机理与 3.5.1 节中给出的线性平滑滤波器一样。这里只是简单地使用了不同的系数。

对角线方向也可以这样组成：在数字拉普拉斯变换的定义中，在式(3.6-6)中添入两项，即两个对角线方向各加一项。每个新添加项的形式与式(3.6-4)或式(3.6-5)类似，只是其坐标轴的方向沿对角线方向。由于对角线方向上的每项还包含一个 $-2f(x,y)$，所以现在从不同方向的项中应总共减去 $-8f(x,y)$。图3.37(b)显示了执行这一新定义的模板。这种模板对 45° 增幅的结果是各向同性的。我们在实践中可能常常见到图 3.37(c)和(d)所示的拉普拉斯模板。它们是由式(3.6-4)和式(3.6-5)中用过的二阶微分的定义得到的，只是其中的 1 是负的。因此，它们产生了等效的结果，但在将拉普拉斯滤波后的图像与其他图像合并(相加或相减)时，必须考虑符号上的差别。

由于拉普拉斯是一种微分算子，因此其应用着重于图像中的灰度突变区域，而非灰度级缓慢变化的区域。这将产生暗色背景中叠加有浅灰色边线和突变点的图像。将原图像和拉普拉斯图像叠加在一起的简单方法，可以复原背景特性并保持拉普拉斯锐化处理的效果。如上一段中所述，使用的拉普拉斯定义很重要。如果所用定义的中心系数是负的，要得到锐化结果，就必须从原图像减去经拉普拉斯变换后的图像，而不是加上它。因此，我们使用拉普拉斯增强图像的基本方法可表示为下式：

182~184

$$g(x,y)=f(x,y)+c\left[\nabla^2 f(x,y)\right] \tag{3.6-7}$$

式中，$f(x,y)$ 和 $g(x,y)$ 分别是输入图像和锐化后的图像。如果使用图3.37(a)或图3.37(b)中的拉普拉斯滤波器，则常数 $c=-1$，如果使用另外两个滤波器，则常数 $c=1$。

a b c d

0	1	0
1	−4	1
0	1	0

1	1	1
1	−8	1
1	1	1

0	−1	0
−1	4	−1
0	−1	0

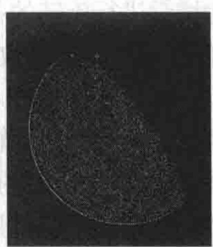
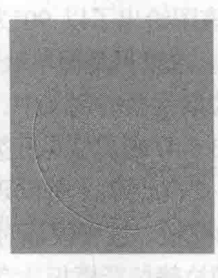
−1	−1	−1
−1	8	−1
−1	−1	−1

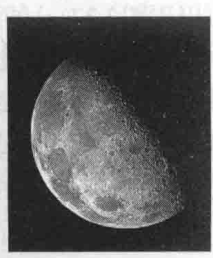
图 3.37 (a)实现式(3.6-6)所用的滤波器模板；(b)用于实现带有对角项的公式的扩展模板；(c)~(d)实践中常用的其他两个拉普拉斯实现

例 3.15 使用拉普拉斯算子锐化图像。

图3.38(a)显示了月球北极的一幅略显模糊的图像。图3.38(b)显示了使用图3.37(b)中的拉普拉斯模板对该图像滤波后的结果。由于拉普拉斯图像中既有正值又有负值，且所有负值在显示时都修剪为 0，所以图像的大部分是黑色的。

标定拉普拉斯图像的一种典型方法是，先让其最小值加上一个代替零的最小值，然后将结果标定到整个灰度范围 $[0, L-1]$ 内，如同式(2.6-10)和式(2.6-11)中解释的那样。图3.38(c)是用这种方法标定过的图像。注意，该图像的支配特征位于边缘和灰度的不连续处。前述的黑色背景由于标定，现在已变成灰色。这种浅灰色的外观是被适当标定的典型拉普拉斯图像。图3.38(d)显示了在 $c=-1$ 时使用式(3.6-7)得到的结果。该图像的细节比原图像更清晰。原图像与拉普拉斯处理结果相加，可以使图像中的各灰度值复原，而且拉普拉斯处理增强了图像中灰度突变处的对比度。最终结果是使图像中的细节部分得到了增强，并很好地保留了图像的背景色调。最后，图 3.38(e)显示了使用图 3.37(b)中的滤波器重复前述过程的结果。这里，我们注意到图3.38(d)的清晰度明显提升。因为使用了图3.37(b)的滤波器，导致在对角方向上产生了额外的区别（锐化），这并不令人意外。像图3.38(d)和图3.38(e)显示的这些结果，使得拉普拉斯算子成为锐化数字图像的一种工具。



a b c d e

图 3.38 (a)月球北极的模糊图像；(b)未标定图像经拉普拉斯滤波后的结果；(c)标定图像经拉普拉斯滤波后的结果；(d)用图 3.37(a)中的模板锐化后的图像；(e)用图 3.37(b)中的模板锐化后的图像(原图像由 NASA 提供)

3.6.3 非锐化掩蔽和高提升滤波

印刷和出版业已用多年的图像锐化处理过程是从原图像中减去一幅非锐化(平滑)的版本。这个称为非锐化掩蔽的处理过程由下列步骤组成：

1. 模糊原图像。
2. 从原图像中减去模糊图像(产生的差值图像称为模板)。
3. 将模板加到原图像上。

184 ~ 185

令 $\bar{f}(x,y)$ 表示模糊图像，非锐化掩蔽以公式形式描述如下。首先，我们得到模板：

$$g_{\text{mask}}(x,y)=f(x,y)-\bar{f}(x,y) \tag{3.6-8}$$

然后，在原图像上加上该模板的一个权重部分：

$$g(x,y)=f(x,y)+k*g_{\text{mask}}(x,y) \tag{3.6-9}$$

通常，我们会在上式中包含一个权重系数 $k(k\geqslant 0)$。当 $k=1$ 时，我们得到上面定义的非锐化掩蔽。当 $k>1$ 时，该处理称为高提升滤波。选择 $k<1$ 则不强调非锐化模板的贡献。

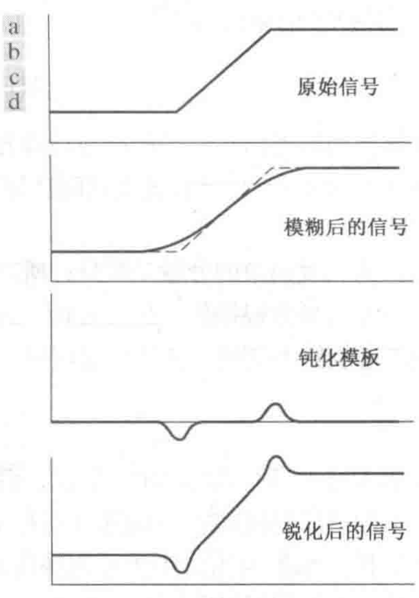


图 3.39　非锐化掩蔽机理的说明：(a) 原始信号；(b) 模糊信号，为便于参考，原始信号使用虚线显示；(c) 非锐化模板；(d) 将图 (c) 和图 (a) 相加得到的锐化后的信号

图 3.39 解释了非锐化掩蔽是如何工作的。图 3.39 (a) 中的灰度剖面图可以解释为通过垂直边缘的水平扫描线，垂直边缘是图像中从暗区到亮区的过渡。图 3.39 (b) 显示了平滑后的结果，为了参考，该结果已叠加在原始信号(显示为虚线)上。图 3.39 (c) 是非锐化模板，它是从原始信号中减去模糊信号得到的。通过将该结果与对应于图 3.36 (a) 中斜坡的图 3.36 (c) 中的部分比较，我们发现，图 3.39 (c) 中的非锐化模板与我们使用二阶微分得到的结果非常相似。图 3.39 (d) 是最后锐化的结果，它是通过把模板加到原始信号上得到的。现在强调(锐化)了信号中出现灰度斜率变化的点。观察添加到原信号中的负值。这样，如果原图像有任何零值，或如果选择的 k 值大到足以使模板峰值大于原信号中的最小值时，那么最终的结果可能会存在负灰度。负值将导致边缘周围出现暗色晕轮，k 足够大时会产生不好的结果。

例 3.16　使用非锐化掩蔽锐化图像。

图 3.40(a) 显示了暗灰背景上带有略微模白色文字的图像。图 3.40 (b) 是使用大小为 5×5 且 $\sigma=3$ 的高斯平滑滤波器得到的结果(见 3.4.4 节)。图 3.40 (c) 是使用式 (3.6-8) 得到的非锐化模板。图 3.40 (d) 是用非锐化模板［式 (3.6-9), $k=1$］得到的结果。该图像较原图像略有改进，但改进应能更为明显。图 3.40 (e) 显示了使用式 (3.6-9) 并令 $k=4.5$ 的结果，这是我们能用的最大可能值，且在最终结果中始终保持所有值为正。该图像与原图像相比已有明显改进。

DIP-XE　DIP-XE　DIP-XE　DIP-XE　DIP-XE　a b c d e

图 3.40　(a) 原图像；(b) 使用高斯滤波器模糊的结果；(c) 非锐化模板；(d) 使用非锐化掩蔽的结果；(e) 使用高提升滤波的结果

3.6.4　使用一阶微分锐化(非线性)图像——梯度

图像处理中的一阶微分是用梯度幅值来实现的。对于函数 $f(x,y)$，f 在坐标 (x,y) 处的梯度定义为二维列向量

$$\nabla f\equiv \text{grad}(f)\equiv\begin{bmatrix}g_x\\ g_y\end{bmatrix}=\begin{bmatrix}\dfrac{\partial f}{\partial x}\\[2mm] \dfrac{\partial f}{\partial y}\end{bmatrix} \tag{3.6-10}$$

10.2.5 节将详细讨论梯度。这里，我们仅对使用梯度幅值锐化图像感兴趣。

185
~
187

该向量具有重要的几何特性，即它指出了在位置(x,y)处f的最大变化率的方向。

向量∇f的幅值(长度)表示为$M(x,y)$，即

$$M(x,y)=\text{mag}(\nabla f)=\sqrt{g_x^2+g_y^2} \tag{3.6-11}$$

它是梯度向量方向变化率在(x,y)处的值。注意，$M(x,y)$是与原图像大小相同的图像，它是当x和y允许在f中的所有像素位置变化时产生的。在实践中，该图像通常称为梯度图像(含义清楚时，可简称为梯度)。

因为梯度向量的分量是微分，所以它们是线性算子。但该向量的幅度不是线性算子，因为求幅度是做平方和平方根操作。另一方面，式(3.6-10)中的偏微分不是旋转不变的(各向同性)，而梯度向量的幅度是旋转不变的。在某些实现中，用绝对值来近似平方和平方根操作更适合于计算：

$$M(x,y)\approx|g_x|+|g_y| \tag{3.6-12}$$

该表达式仍保留了灰度的相对变化，但通常丢失了各向同性特性。然而，像拉普拉斯的情况那样，后面定义的离散梯度的各向同性仅在有限旋转增量的情况下才被保留，它依赖于所用的近似微分的滤波器模板。结果表明，用于近似梯度的常用模板在90°的倍数时是各向同性的。这些结果与我们使用式(3.6-11)还是使用式(3.6-12)无关，因此在我们选择这样做时，使用后一公式对结果并无影响。

如拉普拉斯情况下那样，我们现在离散化前面的公式，并由此形成合适的滤波模板。为简化下面的讨论，我们将使用图3.41(a)中的符号来表示一个3×3区域内图像点的灰度。例如，使用图3.28中引入的符号，令中心点z_5表示任意位置(x,y)处的$f(x,y)$；z_1表示为$f(x-1,y-1)$；等等。如3.6.1节指出的那样，满足3.6.1节所声明条件的对一阶微分的最简近似是$g_x=(z_8-z_5)$和$g_y=(z_6-z_5)$。在早期数字图像处理的研究中，由Roberts[1965]提出的其他两个定义使用交叉差分：

$$g_x=(z_9-z_5)\text{ 和 }g_y=(z_8-z_6) \tag{3.6-13}$$

如果使用式(3.6-11)和式(3.6-13)，则梯度图像计算为

$$M(x,y)=\left[(z_9-z_5)^2+(z_8-z_6)^2\right]^{1/2} \tag{3.6-14}$$

如果用式(3.6-12)和式(3.6-13)，则

$$M(x,y)\approx|z_9-z_5|+|z_8-z_6| \tag{3.6-15}$$

按此前的描述方式,很容易理解x和y会随图像的维数变化。式(3.6-13)中所需的偏微分项可以用图3.41(b)中的两个线性滤波器模板来实现。这些模板称为*罗伯特交叉梯度算子*。

偶数尺寸的模板很难实现，因为它们没有对称中心。我们感兴趣的最小模板是3×3模板。使用以z_5为中心的3×3邻域对g_x和g_y的近似如下式所示：

$$g_x=\frac{\partial f}{\partial x}=(z_7+2z_8+z_9)-(z_1+2z_2+z_3) \tag{3.6-16}$$

和

$$g_y=\frac{\partial f}{\partial y}=(z_3+2z_6+z_9)-(z_1+2z_4+z_7) \tag{3.6-17}$$

这两个公式可以使用图3.41(d)和图3.41(e)中的模板来实现。使用图3.41(d)中的模板实现的3×3图像区域的第三行和第一行的差近似x方向的偏微分，另一个模板中的第三列和第一列的差近似y方向的微分。用这些模板计算偏微分后，就得到了之前所说的梯度幅值。例如，将g_x和g_y代入

187 ~ 189

式(3.6-12)得到

$$M(x,y) \approx |(z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3)| + |(z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)| \tag{3.6-18}$$

图3.41(d)和图3.41(e)中的模板称为 Sobel 算子。中心系数之所以使用权重 2，原因是通过突出中心点的作用而达到平滑的目的(第10章中将详细介绍该思想)。注意，图3.41所示的所有模板中的系数总和为 0，这正如微分算子的期望值那样，表明灰度恒定区域的响应为 0。

a b c d e

(a)

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

(b)

−1	0
0	1

(c)

0	−1
1	0

(d)

−1	−2	−1
0	0	0
1	2	1

(e)

−1	0	1
−2	0	2
−1	0	1

图 3.41　(a)一幅图像的3×3区域(z 是灰度值)；(b)~(c)罗伯特交叉梯度算子；(d)~(e)所有的模板系数之和为零，如微分算子预料的那样

如前所述，g_x 和 g_y 的计算是线性操作，因为它们涉及微分操作，因此可以使用图 3.41 中的空间模板如乘积求和那样实现。使用梯度进行非线性锐化时，涉及 $M(x, y)$ 的平方和平方根计算，或使用绝对值计算代替，所有这些计算都是非线性操作。该操作是在得到 g_x 和 g_y 线性操作后执行的操作。

例 3.17　使用梯度增强边缘。

工业检测中常使用梯度处理，梯度处理一般用于辅助人工检测产品缺陷，更常用于自动检测的预处理。我们将在第 10 章和第 11 章中详细介绍梯度处理。下面考虑一个简单的例子，以展示梯度法如何增强缺陷并消除缓慢变化背景中的特征。在这个例子中，增强用做自动检测的预处理步骤，而不用于人为分析。

图 3.42(a)显示了一幅隐形眼镜的光学图像，它由设计用于突出缺陷的发光装置来照明，例如 4 点方向和 5 点方向眼镜边界上的两个边缘缺陷。图 3.42(b)显示了使用式(3.6-12)及图 3.41(d)与图 3.41(e)中的两个 Sobel 模板得到的梯度图像。在该图像中，边缘缺陷清晰可见，并且还有一个附加的优点，即去掉了灰度不变或变化缓慢的图案阴影，因此简化了自动检测所要求的计算任务。梯度处理还可突出灰度图像中看不见的小斑点(这些小斑点可能是外来物、保护液中的气泡或眼镜中的小缺陷)。在灰度平坦区域中增强小突变的能力，是梯度处理的另一个重要特性。

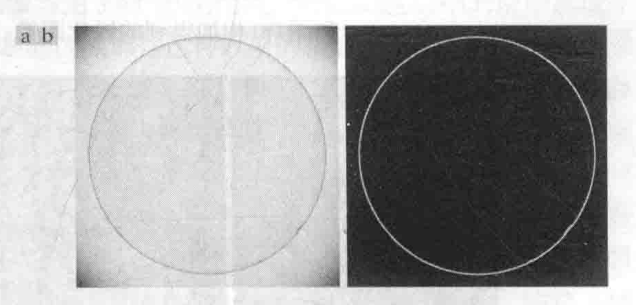


图 3.42　(a)隐形眼镜的光学图像(注意 4 点和 5 点方向边界上的缺陷)；(b)Sobel 梯度(原图像由 Perceptics 公司的 Pete Sites 提供)

3.7　混合空间增强法

到目前为止，除了一些特殊情况，如带有阈值处理(图 3.34)的混合模糊，我们主要关注的还是单一的增强法。通常，为了达到令人满意的结果，对给定的任务需要应用多种互补的图像增强技术。本节将通过一个例子来说明怎样将本章中提到的多种图像增强方法结合起来，以便完成困难的图像增强任务。

189 ~ 191

图3.43(a)所示的图像是一幅人体骨骼核扫描图像，常用于检查人体疾病，如骨骼感染和肿瘤。我们的目的是通过图像锐化突出骨骼的更多细节来增强图像。由于图像灰度的动态范围很窄，并且有很高的噪声，因此很难对其进行增强。对此，我们采取的策略是，首先用拉普拉斯法突出图像中的小细节，然后用梯度法突出其边缘。平滑后的梯度图像将用于掩蔽拉普拉斯图像(关于掩蔽的说明见图 2.30)，稍后会对此简短说明。最后，我们将试图使用灰度变换来增大图像的灰度动态范围。

图 3.43(b)显示了使用图 3.37(d)中的滤波器得到的原图像的拉普拉斯操作结果。该图像使用与图 3.38(c)相同的技术标定过(仅为便于显示)。然后，根据式(3.6-7)，简单地将图3.43(a)与图3.43(b)相加，就可得到一幅锐化后的图像。根据图 3.43(b)中的噪声水平，我们可以预料，如果将图 3.43(a)和图3.43(b)相加，会导致相当多的噪声，这可在图3.43(c)中得到证实。于是我们可立刻想到，降低噪声的一种方法是使用中值滤波器。然而，中值滤波器是一种非线性滤波器，它有可能改变图像的性质。这在医学图像处理中是不能接受的。

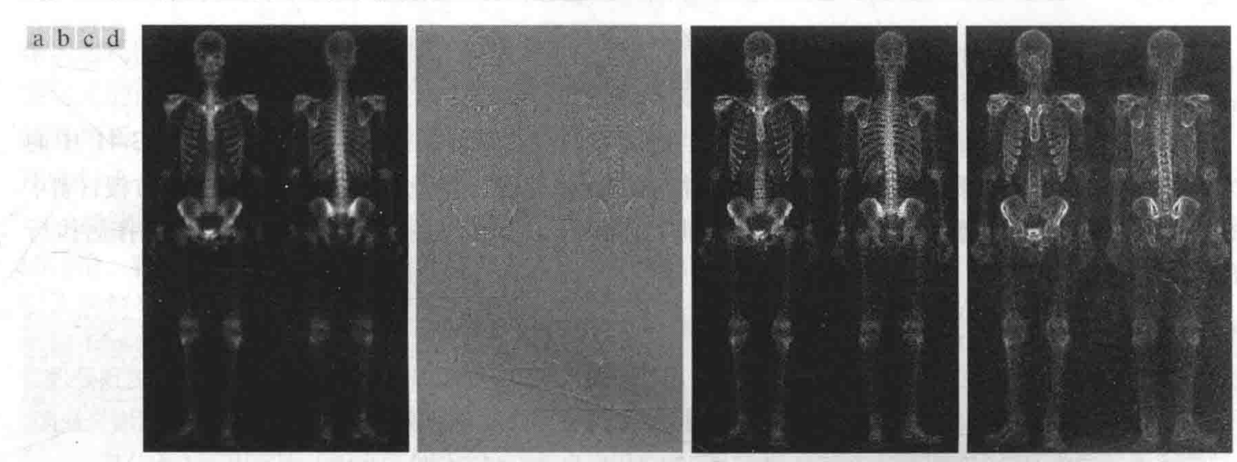


图 3.43 (a)全身骨骼扫描图像；(b)图(a)经拉普拉斯操作后的结果；(c)将图(a)和图(b)相加得到的锐化图像；(d)图(a)经 Sobel 梯度处理后的结果

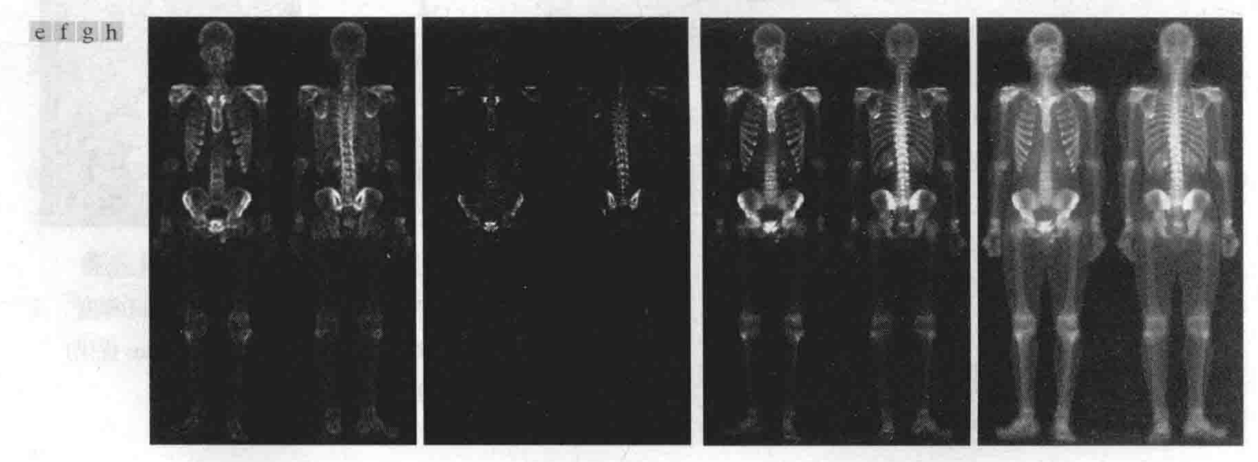


图 3.43(续) (e)使用5×5 均值滤波器平滑后的 Sobel 图像；(f)由图(c)和图(e)相乘形成的掩蔽图像；(g)由图(a)和图(f)求和得到的锐化后的图像；(h)对图(g)应用幂律变换得到的最终结果。请将图(g)与图(a)比较(原图像由 G. E. Medical Systems 公司提供)

191 ~ 193

另一种方法是，使用原图像经梯度操作后的平滑形式所形成的一个模板。这种方法的动机很直接，并且基于 3.6.1 节中介绍过的一阶微分和二阶微分的性质。拉普拉斯操作作为一种二阶微分算子，在图像细节的增强处理方面有一定的优点，且在增强细节方面它是最好的。但与梯度操作相比，拉普拉斯操作会导致更多的

噪声。这些噪声在平滑区域内更令人讨厌，因为在这些区域它非常显眼。梯度变换在灰度变化的区域（灰度斜坡或台阶）的平均响应，要比拉普拉斯操作的平均响应更强烈。梯度操作对噪声和小细节的响应要比拉普拉斯操作的响应弱，而且可以通过均值滤波器对其进行平滑处理来进一步降低。这时，思路是对梯度图像进行平滑处理，并用拉普拉斯图像与它相乘。在这种情况下，可以将平滑后的梯度图像视为一个模板图像。乘积会保留灰度变化强烈区域的细节，同时降低灰度变化相对平坦区域的噪声。我们可以认为，这种处理基本结合了拉普拉斯操作与梯度操作的优点。把结果加到原图像上，就可以得到最终的锐化图像。

图3.43(d)显示了原图像使用式(3.6-12)给出的 Sobel 梯度操作的结果。分量 g_x 和 g_y 分别使用图 3.41(d)和图 3.41(e)中的模板获得。如所期望的那样，该图像中的边缘要比拉普拉斯图像中的边缘突出许多。图 3.43(e)中平滑后的梯度图像是使用大小为 5 × 5 的一个均值滤波器得到的。为便于显示，两幅梯度图像已与拉普拉期图像相同的方式进行了标定。因为梯度图像的最小可能值是 0，故标定后的梯度图像中的背景是黑色的，而不是标定后的拉普拉斯图像中的灰色。图3.43(d)和图3.43(e)比图 3.43(b)明亮的事实再次表明，具有重要边缘内容的梯度图像的值，一般要比拉普拉斯图像的值高。

拉普拉斯图像与平滑后的梯度图像的乘积显示在图3.43(f)中。注意增强了边缘，并降低了噪声，这是使用一幅平滑后的梯度图像来掩蔽拉普拉斯图像的关键目的。将乘积图像与原图像相加就产生了图3.43(g)所示的锐化图像。与原图像相比，该图像中大部分细节的清晰度明显增加，包括肋骨、脊椎骨、骨盆及颅骨。单独使用拉普拉斯操作或梯度操作都不可能达到这种程度的改进。

以上讨论的锐化过程在感知方面并未影响图像灰度的动态范围。因此，增强处理任务的最后一步是增大锐化后图像的动态范围。如 3.2 节和 3.3 节中详细讨论的那样，有很多可以完成这一目标的灰度变换函数。从 3.3.2 节的结果可以知道，对类似此处灰度分布比较暗的图像进行直方图均衡的效果似乎并不好。直方图规定化可能是一种解决方法，但待处理图像的暗特征用幂律变换处理更好。由于我们希望能够扩展灰度范围，因此式(3.2-3)中的γ值必须小于 1。使用该式进行一些实验后，我们得到了图3.43(h)所示的结果，此时$\gamma = 0.5$, $c = 1$。将这幅图像与图 3.43(g)相比较，可以看到图3.43(h)中出现了许多重要的新细节。围绕手腕、手掌、脚踝和脚掌的区域都是这种效果的很好例子。人体的整个骨架结构也很明显，包括手臂骨和腿骨。我们还注意到，人体轮廓及人体组织的清晰度不高，这是由于在扩大灰度动态范围来显现细节的同时，也增大了噪声，但图3.43(h)与原图像相比还是明显改进了视觉效果。

刚才讨论的方法是很有代表性的处理方式，即要达到用单种方法无法实现的效果，需要结合使用多种方法。得到期望结果所用的方法取决于具体的应用。在这个例子中，所示图像类型的最终用户很可能是放射线学者。但也存在超出我们讨论范围的许多原因，如物理学家就不依赖于增强的结果来达到诊断目的。然而，通过突出某些细节作为深入分析原图像或一系列图像的线索，增强后的图像是非常有用的。而在其他领域，增强处理后的结果可能就是最终的“产品”。在印刷业、基于图像的产品检测、法院取证、显微、监视及其许多他领域，增强的主要目的是得到具有较多可见细节的图像。

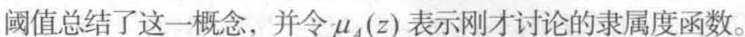
3.8　使用模糊技术进行灰度变换和空间滤波

下面通过介绍模糊集合及其在灰度变换和空间滤波中的应用来结束本章，灰度变换和空间滤波是前几节探讨的主题。业已证明，这两种应用是图像处理中最常使用模糊技术的领域。本章最后的参考文献，提供了关于模糊集合及其在图像处理中的应用的文献。如我们在下面的讨论中看到的那样，模糊集合在解决那些不以精确概念表述的问题时，提供了整个人类知识的框架。

194~195

3.8.1　引言

如 2.6.4 节提及的那样，集合是由一个或多个对象（元素）构成的整体，集合论是处理集合中操作的工具集。与数学逻辑一样，集合论理所当然也是经典数学的基础之一。集合论的核心是集合成员的概念。我们习惯于处理所谓的“干脆”集合，它们的成员在传统的二值布尔逻辑情况下，不是真就是假，用 1 表示真，用 0 表示假。例如，令 Z 表示所有人员的集合，且我们想要定义集合 Z 的一个称为年轻人集合的子集 A。为了形成该子集，我们需要定义一个隶属度函数，它可以对 Z 中的每个成员 z 赋一个 1 值或 0 值。因为我们处理的是二值逻辑，因此隶属度函数简单地定义一个阈值，小于等于该阈值的人考虑为年轻人，大于阈值的人考虑为非年轻人。图 3.44(a)用 20 岁的年龄阈值总结了这一概念，并令 $\mu_A(z)$ 表示刚才讨论的隶属度函数。



隶属度函数也称为特征函数。

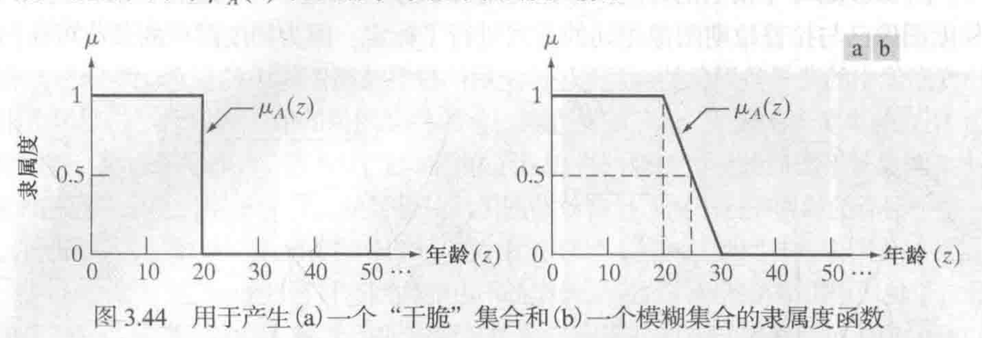


图 3.44　用于产生(a)一个“干脆”集合和(b)一个模糊集合的隶属度函数

使用这一表达，我们会立即面临一个困难：年龄为 20 岁的人被认为是年轻人，但年龄为 20 岁零 1 秒的人就不是年轻人集合的成员。这是使用“干脆”集合的一个基本问题，它在许多实际应用中限制了经典集合论的用途。

我们所需的年轻应是年轻和不年轻之间的逐渐过渡。图 3.44(b)显示了一种可能性。该函数的关键特性是非限制评价，因此能在年轻和不年轻之间连续过渡。这就使得它可能具有“年轻”的程度。现在我们可以说一个人是年轻人(曲线上端的平坦部分)、相对年轻的人(斜坡开始处)、50%年轻的人(斜坡的中间)、不那么年轻的人(斜坡的末端)，等等[注意图 3.44(b)中曲线的斜率引入了什么是“年轻”的更加模糊的概念]。这些模糊的声明与我们不太严密地谈论年龄时是一致的。这样，我们可以用模糊逻辑基础解释非限制评价隶属度函数，并把使用它们生成的集合视为模糊集合。这些概念将在下一节中正式表述。

3.8.2　模糊集合论原理

模糊集合论是由 L. A. Zadeh 在 40 多年前的一篇论文中引入的(Zadeh[1965])。如下面的讨论那样，模糊集合为处理不严密信息提供了一种方式。

为表达给定应用中集合的广泛性，我们采用传统的模糊集合表示法，即集合符号 Z 而不是更为传统的集合符号 U。

定义

令 Z 为元素(对象)集，z 表示 Z 的一类元素，即 $Z = \{z\}$。该集合称为论域。Z 中的模糊集合[①] A 由隶属度函数 $\mu_A(z)$ 表征，它是与 Z 的元素相关的在区间[0, 1]内的一个实数。$\mu_A(z)$ 在 z 处的值表示 A 中 z 的隶属度等级。最接近 $\mu_A(z)$ 的值与在 A 中的 z 的较高隶属度等级趋于一致，而当 $\mu_A(z)$ 接近于 0 时，则相反。“属于”这一概念在普通集合中我们很熟悉，但在模糊集合论中却有不同的含义。对于普通集合，我们说一个元素属于

195 ~ 196

① 文献中也使用术语模糊子集来表示 A 是 Z 的子集，但模糊集合使用得更为频繁。

或不属于一个集合；而对于模糊集合，我们说对于 $\mu_A(z)=1$ 的所有 z 都是集合的完全成员，对于 $\mu_A(z)=0$ 的所有 z 都不是集合的成员，而 $\mu_A(z)$ 的值介于 0 和 1 之间的所有 z 是集合的部分成员。因此，模糊集合是一个由 z 值和(赋予 z 成员等级的)相应隶属度函数组成的序对，即

$$A=\{z,\mu_A(z)\,|\,z\in Z\} \tag{3.8-1}$$

变量连续时，该式中集合 A 有无限个元素。当 z 的值是离散值时，我们可以明确地说明 A 的元素。例如，若图 3.44 中的年龄增量限制为整数年，则有

$$A=\{(1,1),(2,1),(3,1),\cdots,(20,1),(21,0.9),(22,0.8),\cdots,(25,0.5)(24,0.4),\cdots,(29,0.1)\}$$

这里，作为一个例子，元素(22, 0.8)表示年龄 22 在集合中的隶属度等级为 0.8。年龄为 20 岁和低于 20 岁的所有成员都是该集合的完全成员，而年龄为 30 岁和高于 30 岁的成员都不是集合的成员。注意，该集合的曲线是图3.44(b)中曲线上的离散点，所以 $\mu_A(z)$ 完全定义了 A。从另一个角度观察，我们发现一个(离散)模糊集合只不过是一个函数的点的集合，该函数将问题域(论域)中的每个元素映射为一个大于 0 且小于等于 1 的数。这样，我们通常可以交替使用术语模糊集合和隶属度函数。

当 $\mu_A(z)$ 仅有两个值时，例如 0 和 1，隶属度函数就退化成我们熟悉的普通(干脆)集合 A 的特征函数。这样，普通集合就是模糊集合的一种特殊情况。下面，我们考虑涉及模糊集合的几个定义，这些定义是普通集合中相应定义的扩展。

空集：当且仅当 Z 中的隶属度函数等于零，模糊集合为空集。

相等：当且仅当对于所有的 $z\in Z$ 有 $\mu_A(z)=\mu_B(z)$，两个模糊集合 A 和 B 相等。

补集：由 $\bar{A}$ 或 NOT(A)表示模糊集合 A 的补集(NOT)，定义为其隶属度函数是

$$\mu_{\bar{A}}(z)=1-\mu_A(z) \tag{3.8-2}$$

的集合，对于所有的 $z\in Z$。

子集：当且仅当对于所有的 $z\in Z$，有

$$\mu_A(z)\leqslant\mu_B(z) \tag{3.8-3}$$

模糊集合 A 是模糊集合 B 的子集。

并集：对于所有的 $z\in Z$，两个模糊集合 A 和 B 的并集是具有隶属度函数

$$\mu_U(z)=\max\left[\mu_A(z),\,\mu_B(z)\right] \tag{3.8-4}$$

的并集 U(OR)，表示为 $A\cup B$ 或 A OR B。

交集：两个模糊集合 A 和 B 的交集是对于所有的 $z\in Z$，具有隶属度函数

$$\mu_I(z)=\min\left[\mu_A(z),\,\mu_B(z)\right] \tag{3.8-5}$$

的模糊集合 I，表示为 $A\cap B$ 或 A AND B。

注意，在处理模糊集合时，可使用熟悉的符号 NOT, OR 和 AND 来分别表示补、并和交。

例 3.18　模糊集合定义的说明。

图3.45说明了前面的一些定义。图3.45(a)显示了两个集合 A 和 B 的隶属度函数。图3.45(b)显示了集合 A 的补集的隶属度函数，图3.45(c)显示了集合 A 和 B 的并集的隶属度函数，图3.45(d)显示了这两个集合的交集的隶属度函数。注意，这些图形是我们所熟悉的关于普通集合的补集、并集和交集[①]。

196 ~ 198

① 读者可能在文献中遇到过这样的例子，例如两个模糊集合的交集的隶属度函数曲线下的面积用阴影表示，以指出该操作的结果。这源于普通集合操作，是不正确的。在处理模糊集合时，这仅适用于沿着隶属度函数本身的点。

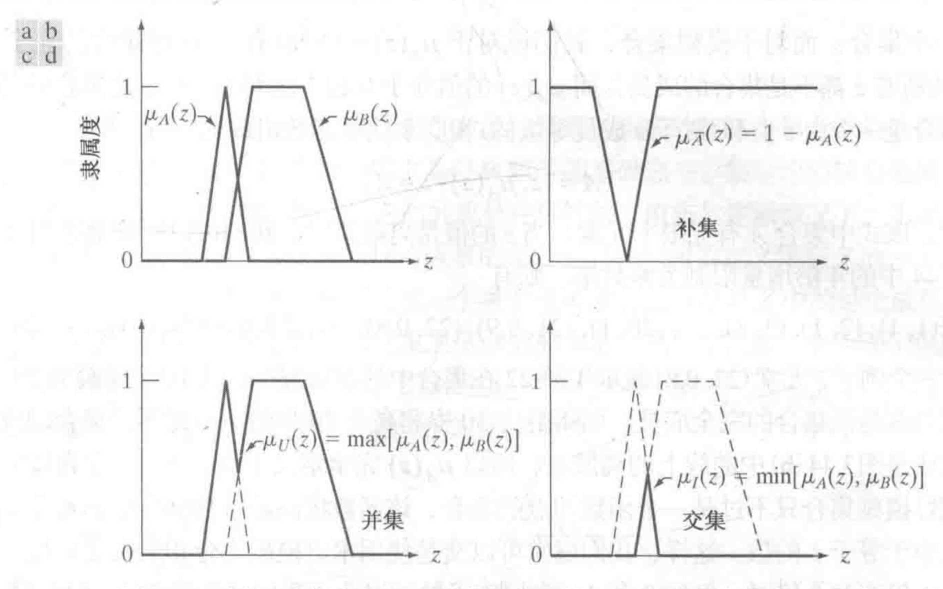


图 3.45　(a)两个集合 A 和 B 的隶属度函数；(b) A 的补集的隶属度函数；(c)~(d)两个集合的并集和交集的隶属度函数

尽管模糊逻辑和概率均在相同的区间[0, 1]上操作，但这两种操作却有明显的区别。考虑图3.44 中的例子，概率的说法可能是“一个人年轻的可能性是 50%”，而模糊逻辑的说法可能是“一个人在年轻人集合中的隶属度等级是 0.5”。这两种说法的区别很重要。在第一种说法中，一个人不是在年轻集合中，就是在不年轻的集合中；我们仅有 50%的机会知道这个人属于哪个集合。第二种说法预设一个人的年轻程度，在这种情况下其年轻程度是 0.5。另一种解释是说这是一个“平均的”年轻人：不是真年轻，但也不是太接近于不年轻。换句话说，模糊逻辑完全不是概率；它仅仅处理一个集合中隶属度的等级。在这种情况下，模糊逻辑这一概念就在由含混和不精确而非随机性表征的应用中找到了用途。

一些常用的隶属度函数

实际中使用的隶属度函数类型如下所示。

三角形：

$$\mu(z)=\begin{cases}1-(a-z)/b, & a-b\leqslant z<a\\ 1-(z-a)/c, & a\leqslant z\leqslant a+c\\ 0, & 其他\end{cases} \tag{3.8-6}$$

梯形：

$$\mu(z)=\begin{cases}1-(a-z)/c, & a-c\leqslant z<a\\ 1, & a\leqslant z<b\\ 1-(z-b)/d, & b\leqslant z\leqslant b+d\\ 0, & 其他\end{cases} \tag{3.8-7}$$

Σ：

$$\mu(z)=\begin{cases}1-(a-z)/b, & a-b\leqslant z\leqslant a\\ 1, & z>a\\ 0, & 其他\end{cases} \tag{3.8-8}$$

198~199

S 形：

$$S(z;a,b,c)=\begin{cases}0, & z<a\\ 2\left(\dfrac{z-a}{c-a}\right)^2, & a\leqslant z\leqslant b\\ 1-2\left(\dfrac{z-c}{c-a}\right)^2, & b<z\leqslant c\\ 1, & z>c\end{cases}\tag{3.8-9}$$

钟形：

$$\mu(z)=\begin{cases}S(z;c-b,c-b/2,c), & z\leqslant c\\ 1-S(z;c,c+b/2,c+b), & z>c\end{cases}\tag{3.8-10}$$

钟形函数有时也称为 Π（或π）函数。

截尾高斯形：

$$\mu(z)=\begin{cases}\mathrm{e}^{-\frac{(z-a)^2}{2b^2}}, & a-c\leqslant z\leqslant a+c\\ 0, & \text{其他}\end{cases}\tag{3.8-11}$$

通常，为简化公式，写成 $\mu(z)$ 的形式时，隶属度函数中仅含自变量 z。为使用式(3.8-10)的形式，我们把式(3.8-9)作为一个例外。图 3.46 显示了刚才讨论的隶属度函数的几个例子。前三个函数是分段线性函数，接着的两个函数是平滑函数，最后一个函数是截尾高斯函数。式(3.8-9)描述了一个重要的 S 形函数，我们在处理模糊集合时会频繁使用这个函数。该式中 $S=0.5$ 时，值 $z=b$ 称为交叉点。如图 3.46(d)所示，这是曲线改变形状的那个点。不难证明 $b=(a+c)/2$（见习题 3.31）。在图 3.46(e)所示的钟形曲线中，b 的值定义为曲线的带宽。

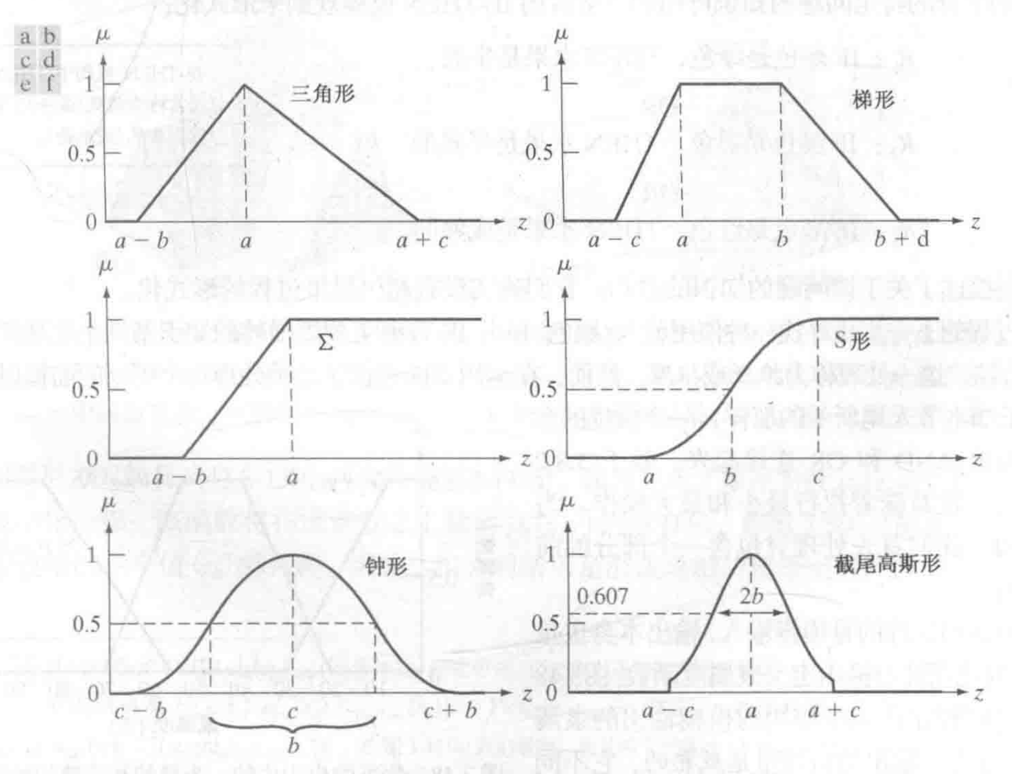


图 3.46　对应于式(3.8-6)至式(3.8-11)的隶属度函数

3.8.3 模糊集合应用

本节关注模糊集合的应用基础，并通过一些简单且熟悉的例子来说明这些概念。然后，我们把这些结果应用到 3.8.4 节和 3.8.5 节的图像处理中。这种说明方法有助于读者更容易理想相关的内容。

假设我们的目的是用颜色把给定的水果分为三类：生的、半熟的和成熟的。假设观察水果成熟的各个阶段得到了如下结论：生的水果是绿色的，半熟的水果是黄色的，成熟的水果是红色的。标志绿、黄和红是颜色感觉的含混描述。首先，这些标志必须以模糊形式来表达，即它们须被模糊化。这是通过将隶属度定义为颜色(光的波长)的函数来实现的，如图3.47(a)所示。在本文中，颜色是一个语言学上的变量，而一种特定的颜色(如有固定波长的红色)是一个语言学上的值。我们用隶属度函数将语言学上的值 z_0 映射到区间[0, 1]来模糊化，如图3.47(b)所示。

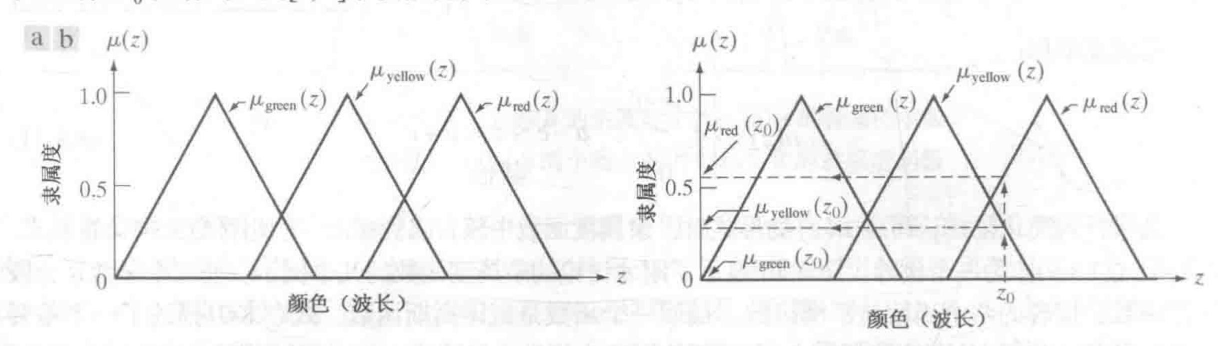


图 3.47 (a)用于模糊颜色的隶属度函数；(b)对特定颜色 z_0 的模糊化(描述颜色感觉的曲线是钟形曲线；见 6.1 节的例子。但在处理模糊集合时，通常使用三角形曲线来近似)

刚刚解释的特定问题的知识可用如下形式的 IF-THEN 模糊规则来形式化：

R_1：IF 颜色是绿色，THEN 水果是生的。

OR

R_2：IF 颜色是黄色，THEN 水果是半熟的。

OR

R_3：IF 颜色是红色，THEN 水果是成熟的。

> IF-THEN 规则中 THEN 左侧的部分通常称为前提(条件)，而右侧的部分称为后果(或结论)。

这些规则描述了关于该问题的知识的总和；它们确实没有超出思维过程的形式化。

该过程的下一步是寻找一种使用输入(颜色)和由 IF-THEN 规则描述的知识基础来创建模糊系统的输出的方法。这一处理称为推断或推理。然而，在运用推断之前，必须处理每个规则的前提以便产生单个值。正如本节末尾所示的那样，一个前提的多个部分可由 AND 和 OR 连接起来。基于 3.8.2 节的定义，这意味着执行最小和最大操作。为简化说明，我们首先处理只包含一个部分的前提的规则。

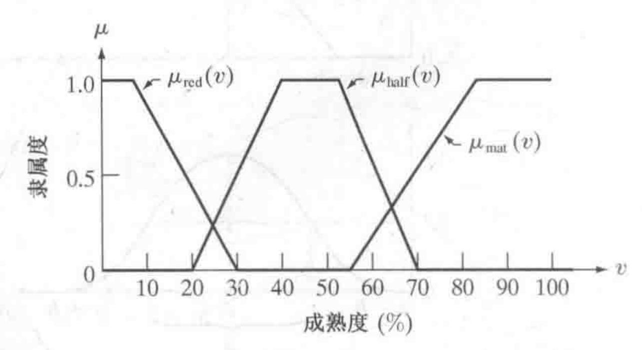


图 3.48 表征输出为生的、半熟的和成熟的的隶属度函数

因为我们处理的是模糊输入，输出本身也是模糊的，因此须针对输出定义隶属度函数。图3.48 显示了我们将在该例中使用的模糊输出的隶属度函数。注意，输出的自变量是成熟的，它不同于输入的自变量。

200 ~ 202

图3.47和图3.48与基础规则一起，包含了所要求的与输入和输出相关的所有信息。例如，我们注意到，“红色 AND 成熟”的表达与前面定义的交集(AND)操作相同。在当前情况下，输入和输出的隶属度函数的自变量是不同的，因此结果将是二维的。例如，图 3.49(a)和图 3.49(b)显示了红色和成熟的隶属度函数,图3.49(c)显示了它们的二维关系。为了得到两个函数间 AND 操作的结果,回忆式(3.8-5)可知，AND 定义为两个隶属度函数的最小，即

$$\mu_3(z, v) = \min\{\mu_{\text{red}}(z), \mu_{\text{mat}}(v)\} \tag{3.8-12}$$

式中，下标 3 表示这是基础知识中规则 R_3 的结果。图3.49(d)显示了 AND 操作的结果[①]。

式(3.8-12)是涉及两个隶属度函数的通解。在实践中，我们感兴趣的是来自一个特定输入的输出结果。令 z_0 表示红色的一个特定值。该输入的响应中的红色分量的隶属等级是简单的标量值 $\mu_{\text{red}}(z_0)$。我们找到与规则 R_3 和这个特定输入对应的输出，即在 $\mu_{\text{red}}(z_0)$ 和 $\mu_3(z,v)$ 间执行 AND 操作，并在 z_0 处求值。如前面表明的那样，AND 操作使用最小操作来实现：

$$Q_3(v) = \min\{\mu_{\text{red}}(z_0), \mu_3(z_0, v)\} \tag{3.8-13}$$

式中，$Q_3(v)$ 表示由规则 R_3 和一个特定输入导致的模糊输出。Q_3 中的唯一变量是所期望的输出变量 v。

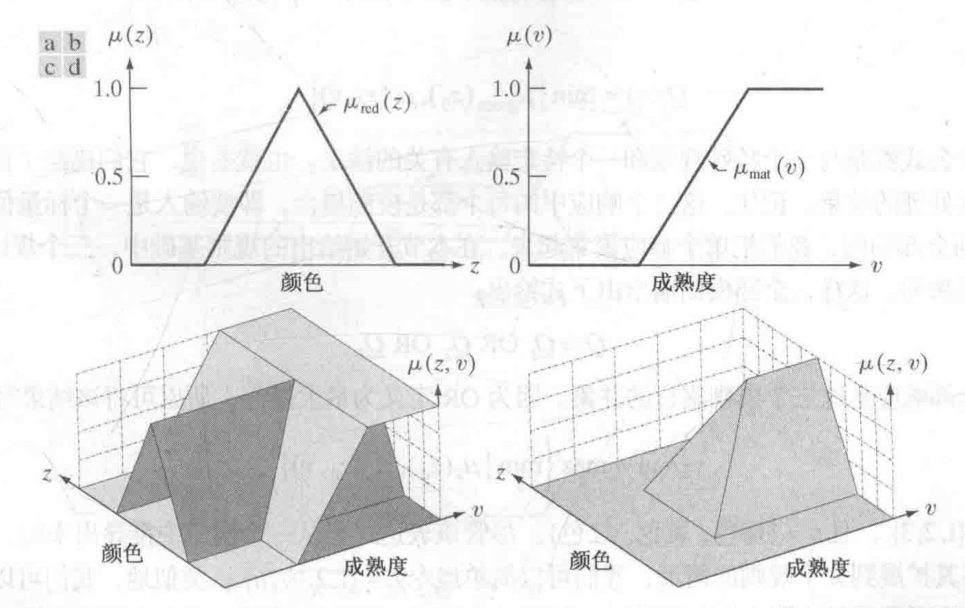


图 3.49　(a)与红色相关联的隶属度函数的形状；(b)相应的输出隶属度函数，两个函数由规则 R_3 相关联；(c)两个函数组合在一起的表示。该表示是二维的，因为在图(a)和图(b)中的自变量是不同的；(d)图(a)和图(b)的 AND 操作结果，如式(3.8-5)的定义一样

为了用图形方式解释式(3.8-13)，再次考虑图3.49(d)，该图显示了通用函数 $\mu_3(z,v)$。执行一个正常数 c 的最小化操作，该函数将在该常数之上裁剪 $\mu_3(z,v)$ 的所有值，如图 3.50(a)所示。然而，我们只对沿彩色轴的一个值(z_0)感兴趣，因此，相关的结果是沿成熟轴的截短函数的剖面，该剖面位

① 注意，式(3.8-12)是由序对值$\{\mu_{\text{red}}(z), \mu_{\text{mat}}(v)\}$形成的，回忆可知，序对集合通常称为笛卡儿积，由 $X\times V$ 表示，其中 X 是由 $\mu_{\text{red}}(z)$ 改变 z 生成的值集 $\{\mu_{\text{red}}(z_1), \mu_{\text{red}}(z_2), \cdots, \mu_{\text{red}}(z_n)\}$，$V$ 是由 $\mu_{\text{mat}}(v)$ 改变 v 生成的类似值集。这样，$X\times V = \{\mu_{\text{red}}(z_1), \mu_{\text{mat}}(v_1), \cdots, (\mu_{\text{red}}(z_n), \mu_{\text{mat}}(v_n))\}$，由图 3.49(d)我们看到，涉及两个变量的 AND 操作可表达为从 $X\times V$ 到区间[0, 1]的映射，表示为 $X\times V \rightarrow [0,1]$。尽管我们在当前的讨论不使用这种表示，但这里提及它的原因在于我们可能会在关于模糊集合的文献中遇到它。

202 ~ 203

于 z_0 处，如图3.50(b)所示［因为图3.50(a)对应于规则 R_3，它满足 $c=\mu_{\text{red}}(z_0)$］。式(3.8-13)是该剖面的表达式。

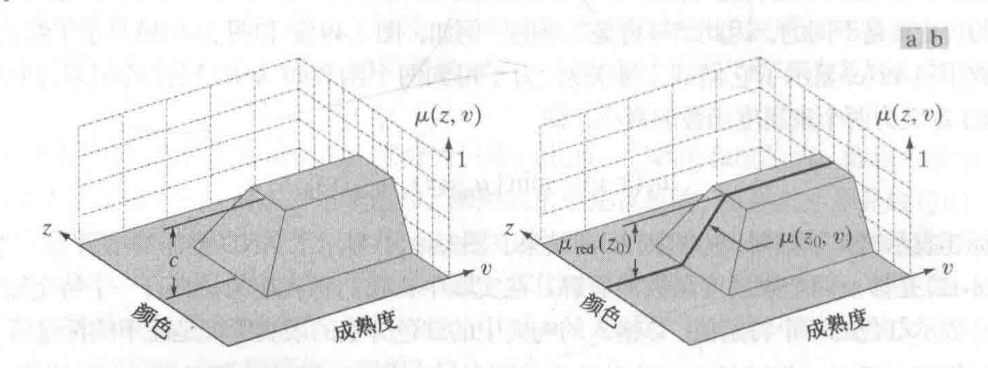


图 3.50 (a)一个任意常数 c 和来自式(3.8-12)的函数 $\mu_3(z,v)$ 的最小化结果。最小化等价于一个 AND 操作；(b)在特定颜色 z_0 处的剖面(黑线)

利用同样的推理，我们可得到由其他两个规则和特定输入 z_0 导致的模糊响应，如下所示：

$$Q_2(v)=\min\left\{\mu_{\text{yellow}}(z_0),\ \mu_2(z_0,\ v)\right\} \tag{3.8-14}$$

和

$$Q_1(v)=\min\left\{\mu_{\text{green}}(z_0),\mu_1(z_0,v)\right\} \tag{3.8-15}$$

上面的每个公式都是与一个特殊规则和一个特定输入有关的输出。也就是说，它们描绘了前面几段中提到的隐含处理的结果。记住，这三个响应中的每个都是模糊集合，即使输入是一个标量值。

为得到全部响应，我们把单个响应聚集起来。在本节开始给出的规则基础中，三个规则是由 OR 操作联系起来的。这样，全部模糊输出由下式给出：

$$Q=Q_1\ \text{OR}\ Q_2\ \text{OR}\ Q_3 \tag{3.8-16}$$

我们看到全部响应是这三个模糊集合的并集。因为 OR 定义为最大操作，所以可将该结果写为

$$Q(v)=\max_r\left\{\min_s\left\{\mu_s(z_0),\ \mu_r(z_0,\ v)\right\}\right\} \tag{3.8-17}$$

式中，$r=\{1,2,3\}$，且 $s=\{$绿色, 黄色, 红色$\}$。尽管该表达式是从一个例子中推导出来的，但它完全通用；要将其扩展到 n 个规则的情形，我们可以简单地令 $r=\{1,2,\cdots,n\}$；类似地，我们可以把 s 扩展到包括任何有限数量的隶属度函数的情形。式(3.8-16)和式(3.8-17)描述了相同的事情：模糊系统的响应 Q，是由该隐含处理每个规则得到的单个模糊集合的并集。

图3.51以图示方式总结了迄今为止讨论的内容。图3.51(a)显示了在 z_0 处计算的三个输入隶属度函数，图 3.51(b)显示了对输入 z_0 的响应的输出。这些模糊集合是使用图3.50(b)的方法裁剪后的剖面。注意，Q_1 由所有的 0 组成，因为 $\mu_{\text{green}}(z_0)=0$；也就是说，$Q_1$ 是空集，正如在 3.8.2 节定义的那样。图 3.51(c)显示了最终结果 Q，它本身是一个由 Q_1，Q_2 和 Q_3 的并集形成的模糊集合。

我们已经成功地得到了相应于特定输入的完整输出，但我们仍然在处理一个模糊集合。最后一步是使用称为去模糊的适当处理，从模糊集合 Q 中得到一个“干脆”输出 v_0。对 Q 去模糊得到“干脆”输出有几种方法。用得最多的一种方法是计算该集合的重心(本章末尾给出的参考文献中讨论了其他方法)。这样，若来自式(3.8-17)的 $Q(v)$ 有 K 种可能的值 $Q(1),Q(2),\cdots,Q(K)$，则其重心由下式给出：

203
~
204

$$v_0 = \frac{\sum_{v=1}^{K} vQ(v)}{\sum_{v=1}^{K} Q(v)} \tag{3.8-18}$$

在图3.51(c)中，用 Q 的离散值[①]计算该公式，得到 $v_0 = 72.3$，它指出给定的颜色 z_0 意味着一种水果的成熟度约为 72%。

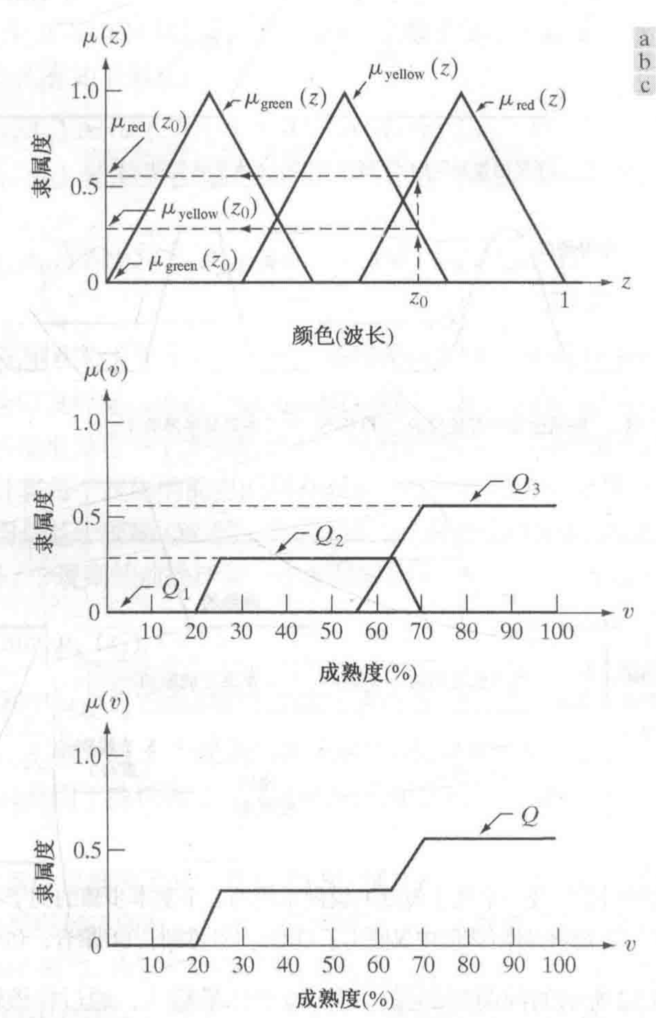


图 3.51　(a)所选特定颜色 z_0 的隶属度函数；(b)由式(3.8-13)到式(3.8-15)得到的单独的模糊集合；(c)用式(3.8-16)或式(3.8-17)得到的最终模糊集合

迄今为止，我们已讨论了 IF-THEN 规则，这些规则的前提只有一部分，例如“如果颜色是红色”。包含更多部分的规则必须组合起来，以便得出描述该规则的整个前提的单个数字。例如，假设我们有一个规则：IF 颜色是红色 OR 坚固性是软的，THEN 水果是成熟的。必须为语言学变量“软的”定义一个隶属度函数。然后，为获得考虑前提的两部分的规则的单个数字，首先要使用“红色”隶属度函数来评估一种给定的“红色”输入值，然后使用“软的”隶属度函数来评估一种给定的“坚固性”值。因为两部分由 OR 连接起来，我们使用两个结果值的最大值[②]。然后，该值用于“干脆”

204～205

① 为清楚起见，图 3.51(c)中的模糊集合 Q 显示为实线，但请记住，我们在本书中处理的是数字量，因此 Q 是一个数字函数。

② 使用 AND 将各部分连接起来的前提条件可使用最小操作进行类似的计算。

成熟输出隶属度函数的推断处理，该函数是与规则相联系的函数。其他过程与以前相同，就像下面的总结中说明的那样。

图3.52显示了使用两个输入“颜色和坚固性”的水果例子。我们可使用该图和前述内容，把基于规则的模糊逻辑应用中的主要步骤总结如下。

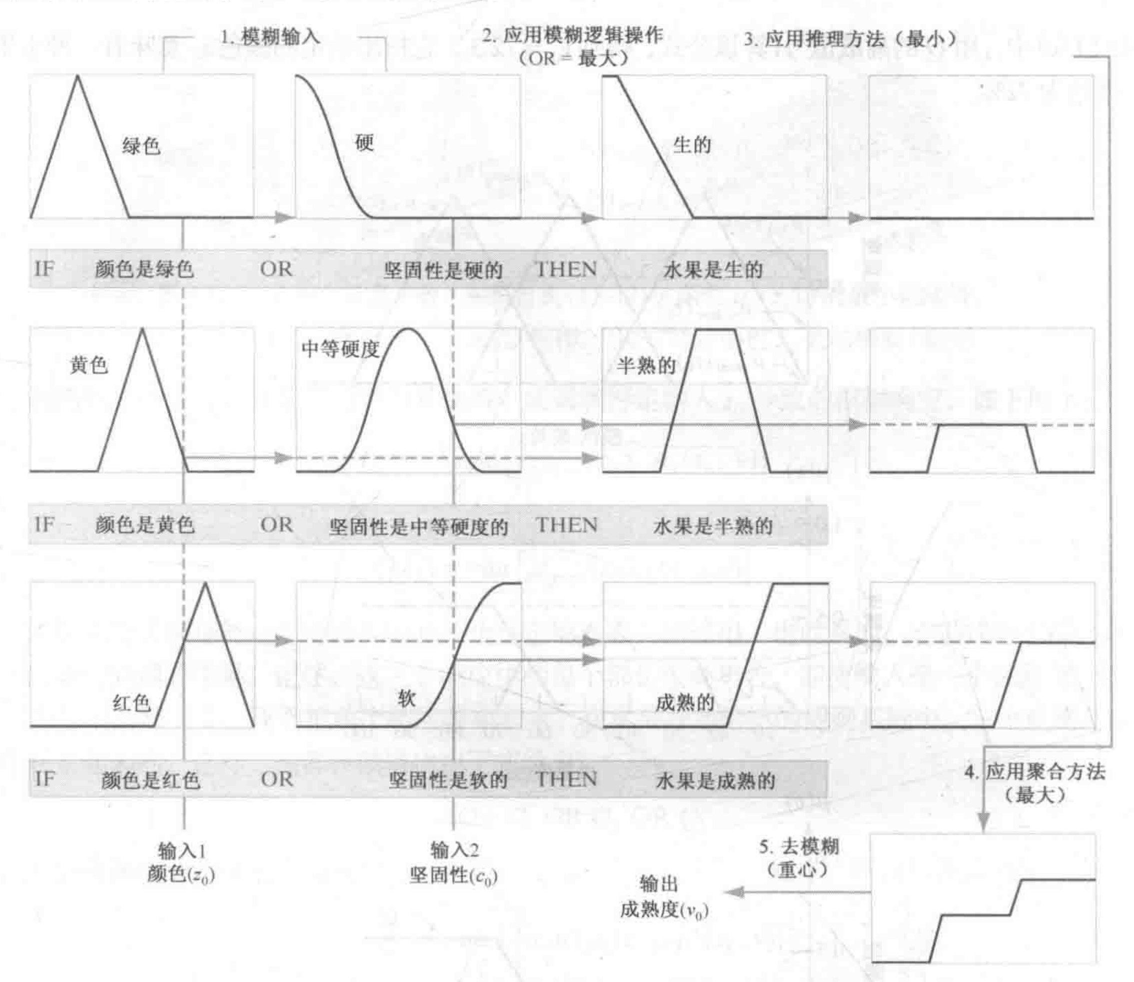


图3.52　说明用于实现一个基于规则的模糊系统的五个基本步骤的例子：(1)模糊；(2)逻辑操作(该例中仅使用了OR)；(3)推断；(4)聚合；(5)去模糊

1. 模糊输入：如图3.52所示的前两列那样。对于每个标量输入，通过将该输入映射到区间[0, 1]，在每个规则中使用可用的隶属度函数，找到相应的模糊值。
2. 执行任何需要的模糊逻辑操作：必须合并前提各部分的输出，并使用最大或最小操作生成单个值，具体是用最大操作还是用最小操作，取决于各部分是由OR连接还是由AND连接。在图3.52中，前提的所有部分由OR连接，因此全部使用最大操作。前提各部分的数量和用于连接这些不同部分的逻辑算子的类型，会因规则的不同而不同。
3. 应用一种推断方法：每个规则的前提的单个输出用来提供与规则相对应的输出。对于推断，我们用AND，它定义为最小操作。这就以这一前提提供的值“修剪”了相应输出的隶属度函数，如图3.52中的第三列和第四列所示。
4. 对来自步骤3的模糊集合应用一种聚合方法：如图3.52中的最后一列所示，每个规则的输出都是一个模糊集合。必须组合这些模糊集合来产生单个输出模糊集合。这里所用的方法是对各个输出使用OR操作，因此采用了最大操作。

206
~
207

5. 对最后输出的模糊集合去模糊：在最后一步中，我们得到了一个“干脆”标量输出。这是通过计算步骤 4 中得到的聚合模糊集合的重心得到的。

当变量的个数很多时，实践中通常使用方便的表示方法(变量，模糊集合)来将一个变量与其对应的隶属度函数配对。例如，规则 IF *颜色是绿色*，THEN *水果是生的*，在这里可以写为 IF $(z$, 绿色$)$ THEN $(v$, 生的$)$，如以前那样，变量 z 和 v 分别表示颜色和成熟度，而绿色和生的分别是由隶属度函数 $\mu_{\text{green}}(z)$ 和 $\mu_{\text{verd}}(v)$ 定义的两个模糊集合。

通常，在处理 M 个 IF-THEN 规则时，N 个输入变量 $z_1, z_2, \cdots, z_N$ 和一个输出变量 v，在图像处理中最常用的模糊规则公式有如下形式：

$$\begin{aligned}
&\text{IF}(z_1, A_{11}) \text{ AND } (z_2, A_{12}) \text{ AND} \cdots \text{AND } (z_N, A_{1N}) \text{ THEN } (v, B_1) \\
&\text{IF}(z_1, A_{21}) \text{ AND } (z_2, A_{22}) \text{ AND} \cdots \text{AND } (z_N, A_{2N}) \text{ THEN } (v, B_2) \\
&\qquad\qquad\cdots \\
&\text{IF}(z_1, A_{M1}) \text{ AND } (z_2, A_{M2}) \text{ AND} \cdots \text{AND } (z_N, A_{MN}) \text{ THEN } (v, B_M) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{ELSE } (v, B_E)
\end{aligned} \tag{3.8-19}$$

式中，A_{ij} 是与第 i 个规则和第 j 个输入变量相关联的模糊集合，B_i 是与第 i 个规则的输出相关联的模糊集合，并且我们已经假设规则的前提分量由 AND 连接。注意，我们已介绍过与模糊集合 B_E 相关联的 ELSE 规则。当前述规则没有一个完全满足时，执行该规则；其输出后面再解释。

如前所述，必须计算每个规则的前提的所有元素，以便得到单个标量值。在图 3.52 中，我们使用了最大操作，因为规则是基于模糊 OR 的。在式(3.8-19)中使用 AND，因此我们必须用最小操作。在式(3.8-19)中，计算第 i 个规则的前提产生一个标量输出 λ_i，它由下式给出：

$$\lambda_i = \min\left\{\mu_{A_{ij}}(z_j); j = 1, 2, \cdots, N\right\} \tag{3.8-20}$$

对于 $i = 1, 2, \cdots, M$，其中 $\mu_{A_{ij}}(z_j)$ 是模糊集合 A_{ij} 在第 j 个输入值计算的隶属度函数。通常，λ_i 称为第 i 个规则的强度水平(或点火水平)。根据前面的讨论，λ_i 只是用于剪切第 i 个规则的输出函数的一个简单的值。

> 在规则集中是使用 OR 还是使用 AND，取决于规则被声明的方式，而声明规则的方式则取决于手边的问题。为便于读者熟悉两者的用法，我们在图 3.52 中使用了 OR，而在式(3.8-19)中使用了 AND。

当 THEN 规则的条件很难满足时，执行 ELSE 规则(3.8.5 节将给出一个如何使用 ELSE 规则的详细例子)。当所有其他规则都很弱时，其响应应该较强。在某种意义上，可以将 ELSE 规则视为在其他规则的结果上执行 NOT 操作。由 3.8.2 节我们知道 $\mu_{\text{NOT}(A)}(z) = \mu_{\bar{A}}(z) = 1 - \mu_A(z)$。然后，在合并(AND)所有级别的 THEN 规则中利用这一概念，给出 ELSE 规则的如下强度水平：

$$\lambda_E = \min\{1 - \lambda_i; i = 1, 2, \cdots, M\} \tag{3.8-21}$$

我们看到，如果所有 THEN 规则在“最大强度”(其响应为 1)处点火，则 ELSE 规则的响应如所期望的那样是 0。THEN 规则的响应较弱时，ELSE 规则的强度增大。这是软件编程中使用的 IF-THEN-ELSE 规则的模糊规则的等效。

在前提中处理 OR 时，我们可以简单地在式(3.8-19)中用 OR 代替 AND，并在式(3.8-20)中用最大(max)代替最小(min)；式(3.8-21)保持不变。尽管可以用公式明确表达更为复杂的前提和由此导致的结果，但仅使用 AND 或 OR 开发的公式是相当通用的，并且广泛应用于各种图像处理应用中。本章末尾的参考文献中包含了其他(但很少使用)的模糊逻辑算子的定义，并探讨了推断(包括多个输出)和去模糊的其他方法。本节前面的介绍是深入了解这一主题的基础。下两节将说明如何把模糊概念用于图像处理。

207 ~ 208

3.8.4 使用模糊集合进行灰度变换

考虑对比度增强这样的一般问题，这是灰度变换的主要应用之一。我们可以用如下规则说明灰度级图像对比度增强的过程：

IF 一个像素是暗的，THEN 使它较暗。

IF 一个像素是灰的，THEN 使它仍是灰的。

IF 一个像素是亮的，THEN 使它较亮。

记住，它们是模糊项，我们可以用图3.53(a)中的隶属度函数来表示暗、灰和亮这些概念。

在输出项中，我们可以将“较暗”考虑为暗灰度值的程度(100%的黑是暗色调的极限值)，将“较亮”考虑为亮色调的程度(100%的白是极限值)，将“灰的”视为中间灰度级中灰度的程度。这里，所谓的“程度”是一种特定灰度的数量。例如，80%的黑是非常暗的灰色。如图3.53(b)所示，当以恒定灰度来解释被修改的强度时，输出隶属度函数是单一值(隶属度函数是常数)。在区间[0, 1]内，当单一值被来自相应规则的响应强度裁剪时，会出现各种程度的灰度，如图3.52 中第四列所示的那样(但应记住，如图中所示，这里仅有一个而非两个输入)。因为在输出隶属度函数中我们处理的是常数，故它满足式(3.8-18)，即对于任何输入 z_0，输出 v_0 由下式给出：

$$v_0 = \frac{\mu_{\text{dark}}(z_0) \times v_d + \mu_{\text{gray}}(z_0) \times v_g + \mu_{\text{bright}}(z_0) \times v_b}{\mu_{\text{dark}}(z_0) + \mu_{\text{gray}}(z_0) + \mu_{\text{bright}}(z_0)} \tag{3.8-22}$$

该式中分子和分母的和比式(3.8-18)中的简单，因为输出隶属度函数是被模糊值修改(裁剪)过的常数。

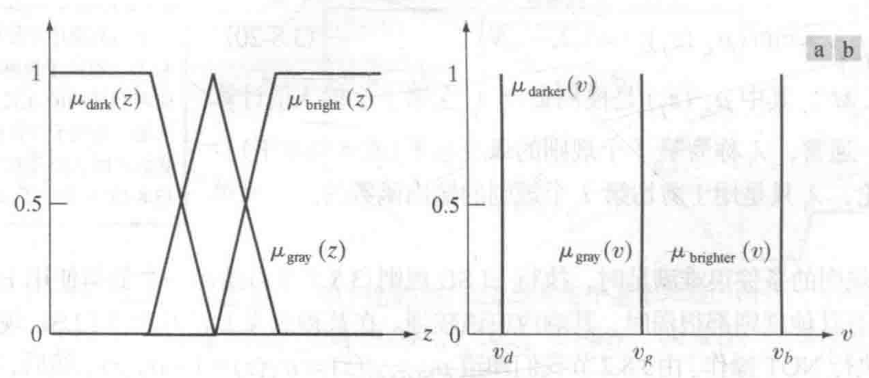


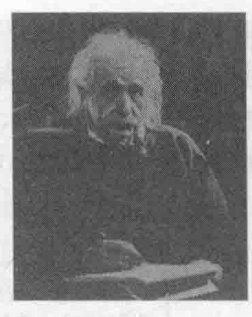
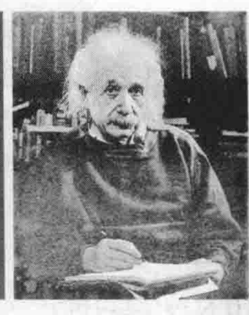
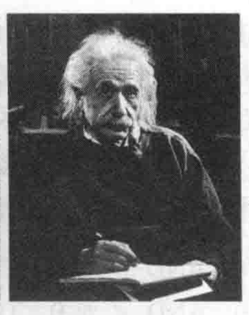
图 3.53 基于规则的对比度增强：(a)输入的隶属度函数；(b)输出的隶属度函数

模糊图像处理的计算量很大，因为全部模糊处理、所有规则的前提、推断、聚集和去模糊，都必须应用到输入图像中的每个像素上。这样，使用像式(3.8-22)中那样的单一值，通过简化推断、聚集和去模糊，可有效地减少计算需求。在处理速度最为重要的应用中，计算量的节省很有意义。

例 3.19 使用基于规则的模糊集的对比度修改的图像增强的说明。

图3.45(a)显示了一幅图像，由于该图像的灰度只占灰度级中很窄的范围[见图3.55(a)中的直方图]，给出的是一幅低对比度外观的图像。作为对比的基础，图3.54(b)是直方图均衡后的结果。正如该结果的直方图显示的那样[见图3.55(b)]，扩展整个灰度级的确增加了对比度，但在高端和低端引入了使图像“过度感光”的外观。例如，爱因斯坦教授的前额和头发的细节大部分都缺失了。图3.54(c)显示了使用前几段讨论的基于规则的对比度修改方法的结果。图3.55(c)显示了所用的叠加在原图像的直方图上的输入隶属度函数。输出单一值选择为：$v_d = 0$ (黑)，$v_g = 127$ (中间灰度)和 $v_b = 255$ (白)。

208 ~ 211

a b c

图 3.54　(a) 低对比度图像；(b) 直方图均衡后的结果；(c) 用基于规则的模糊集合的对比度增强后的结果

比较图 3.54 (b) 和图 3.54 (c)，我们看到后者在色调上有相当大的改进。例如，注意前额和头发中的细节与图3.54 (b) 中相同区域的对比。通过研究示于图3.55 (d) 的 3.54 (c) 的直方图，可很容易地解释这种改进的原因。与均衡后图像的直方图不同，该直方图保留了原图像直方图的基本特征。然而，暗级别(直方图低端的几个高峰值) 的左移很明显，从而暗化了级别。而对于亮级别而言，情形正好相反。中间灰度稍微地分离了一些，但与直方图均衡后的结果相比，分离程度要小很多。

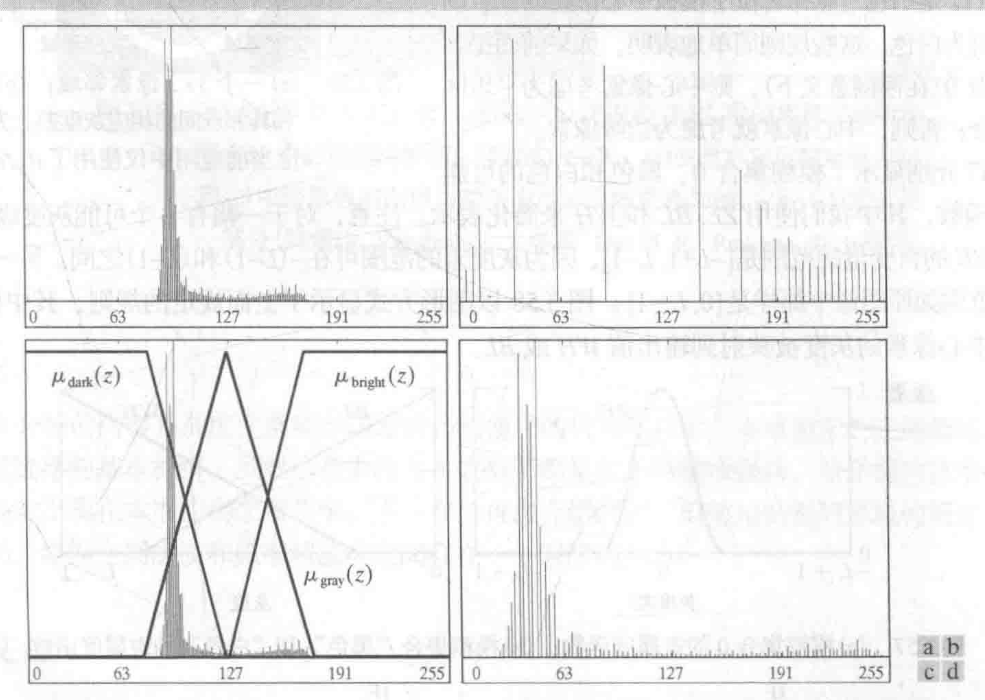


图 3.55　(a)~(b) 图 3.54 (a) 和 (b) 的直方图；(c) 叠加在图 (a) 上的输入隶属度函数；(d) 图 3.54 (c) 的直方图

性能上的这种改进的代价是，加大了处理的复杂性。当处理速度和图像吞吐量是重点考虑的问题时，可采用的实用方法是使用模糊集技术来决定什么是看似能很好地平衡图像的直方图。然后，使用更快的技术如直方图规定化，通过把输入图像的直方图映射到一个或多个用模糊方法确定的“理想”直方图，来实现类似的结果。

3.8.5　使用模糊集合进行空间滤波

把模糊集合用于空间滤波时，基本方法是定义一个邻域特性，该特性“截获”滤波器支持检测的本质。例如，考虑检测一幅图像中各区域间的边缘。这在图像处理的应用中很重要，例如本节前面讨论的锐化和第 10 章将要讨论的图像分割等。

我们可以开发一个基于模糊集合概念的边缘提取算法：“如果一个像素属于平滑区，则令其为白

210 ~ 211

色，否则令其为黑色”，其中“黑色”和“白色”是模糊集合。为了用模糊术语表示“平滑区”这一概念，我们可以考虑邻域中心处像素和邻域像素间的灰度差。对于图3.56(a)中的3×3邻域，中心像素(标为 z_5)和邻域的每个像素间的灰度差形成了图3.56(b)所示的3×3子图像，其中 d_i 表示第 i 个邻点和中心点间的灰度差(即 $d_i = z_i - z_5$，其中 z_5 为灰度值)。4个IF-THEN规则和1个ELSE规则的简单集合就可以实现本段开始提到的模糊集合概念的本质：

IF d_2 是0 AND d_6 是0 THEN z_5 是白色
IF d_6 是0 AND d_8 是0 THEN z_5 是白色
IF d_8 是0 AND d_4 是0 THEN z_5 是白色
IF d_4 是0 AND d_2 是0 THEN z_5 是白色
ELSE z_5 是黑色

> 为简化这个例子，我们仅使用了4邻域和中心点的灰度差。使用8邻域可直接扩展此处显示的方法。

其中，“0”也是一个模糊集合。每个规则的后项定义一个值，中心像素(z_5)的灰度被映射到该值上。也就是说，语句“THEN z_5 是白色”意味着位于模板中心位置的像素的灰度被映射为白色。这些规则简单地表明，如果前面提到的灰度差为0(在模糊意义下)，则中心像素考虑为平坦区域的一部分；否则，中心像素就考虑为边缘像素。

a b

z_1	z_2	z_3
z_4	z_5	z_6
z_7	z_8	z_9

像素邻域

d_1	d_2	d_3
d_4	0	d_6
d_7	d_8	d_9

灰度邻域

图3.56 (a)一个3×3像素邻域；(b)中心像素和其邻点间的相应灰度差。为简化，讨论当前应用中仅使用了 d_2, d_4, d_6 和 d_8

图3.57分别显示了模糊集合0、黑色和白色的可能的隶属度函数，其中我们使用 ZE, BL 和 WH 来简化表示。注意，对于一幅有 L 个可能灰度级的图像，模糊集合 ZE 的自变量的范围是$[-L+1, L-1]$，因为灰度差的范围可在$-(L-1)$和$(L-1)$之间。另一方面，输出灰度的范围如原图像中那样是$[0, L-1]$。图3.58以图形方式显示了上面规定的规则，其中标为 z_5 的方框指出中心像素的灰度被映射到输出值 WH 或 BL。

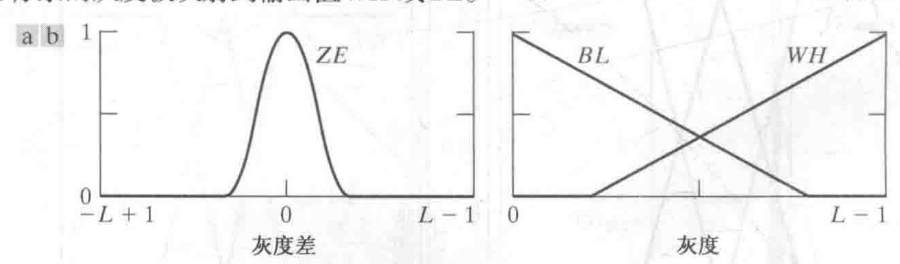


图3.57 (a)模糊集合0的隶属度函数；(b)模糊集合“黑色”和“白色”的隶属度函数

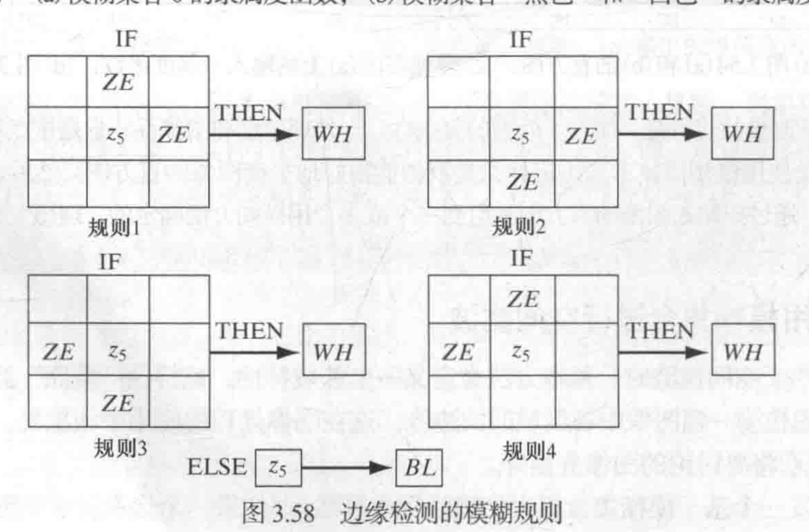
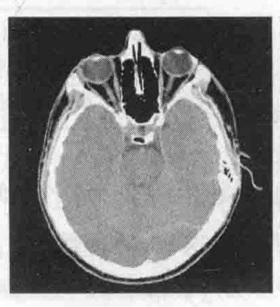
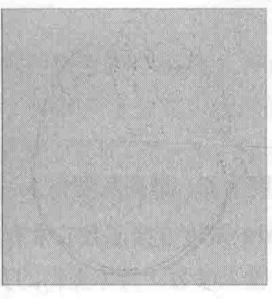


图3.58 边缘检测的模糊规则

211~213

例 3.20　使用基于规则的模糊集合的空间滤波增强边缘的说明。

图3.59(a)显示了一幅大小为 512×512 的人头 CT 扫描图像，图3.59(b)是使用刚刚讨论的模糊空间滤波方法的结果。注意提取区域间边缘的方法的效果，包括大脑的轮廓(灰色区域内部)。图像中的恒定区域呈现灰色，因为当前面讨论的灰度差接近 0 时，THEN 规则有一个很强的响应。这些响应依次裁剪函数 WH。输出(裁剪过的三角形区域的重心)是$(L-1)/2$ 和$(L-1)$间的一个常数，从而在图像中产生了所看到的浅灰色调。该图像的对比度可以通过扩展灰度级来明显改进。例如，图 3.59(c)是通过在 $K = L-1$ 时执行式(2.6-10)和式(2.6-11)中定义的灰度标定得到的。最终结果是图 3.59(c)中灰度值跨过从 0 到$(L-1)$的全部灰度级。


a b c

图3.59　(a)人的头部的 CT 扫描图像；(b)使用图 3.57 中的隶属度函数和图 3.58 中的规则的模糊空间滤波的结果；(c)灰度标定后的结果。(b)和(c)中淡黑色的图像边缘是为增加清晰度添加的，它们不是数据的一部分(原像由 Vanderbilt 大学的 David R. Pickens 博士提供)

小结

本章介绍的内容是灰度变换和空间滤波当前使用的代表性技术。本章包含的主题都是根据领域发展的需求选择的基本材料。虽然本章中的大多数例子都是关于图像增强的，但介绍的技术都很通用，它们会经常出现在本书的剩余章节中。下一章会再次介绍滤波，但使用的是频率域的概念。我们很快就会看到，线性空间滤波和频率域滤波之间存在一一对应的关系。

参考文献

3.1 节中的内容来自 Gonzalez[1986]。3.2 节内容的辅助读物见 Schowengerdt[1983]，Poyton[1996]和 Russ[1999]。关于图像最佳显示的内容，请参阅 Tsujii et al.[1998]。关于直方图处理的早期参考文献有 Hummel[1974]，Gonzalez and Fittes[1977]和 Woods and Gonzalez[1981]。Stark[2000]详细小节了自适应对比度增强的直方图均衡。对比度增强的其他方法见 Centeno and Haertel[1997]和 Cheng and Xu[2000]。严谨直方图规定化的深入阅读材料见 Coltuc, Bolon and Chassery[2006]。局部直方图均衡方法的延伸内容见 Caselles et al.[1999]和 Zhu et al.[1999]。关于图像处理中局部统计应用和实现的内容，可以参考 Narendra and Fitch[1981]。Kim et al.[1997]介绍了一种结合梯度处理与局部统计处理来增强图像的有趣方法。

关于线性空间滤波器及其实现的其他读物有 Umbaugh[2005], Jain[1989]和 Rosenfeld and Kak[1982]。这些文献中还讨论了排序滤波器。Wilburn[1998]也简要介绍了排序滤波器。Pitas and Venetsanopoulos[1990]中包含了中值滤波器和其他非线性滤波器。*IEEE Transaction in Image Processing*[1996]的一期特刊中介绍了非线性图

212 ~ 214

像处理的内容。有关高提升滤波的内容见 Schowengerdt[1983]。在讨论图像复原(见第 5 章)和边缘检测(见第 10 章)时，还会遇到本章中介绍的许多空间滤波器。

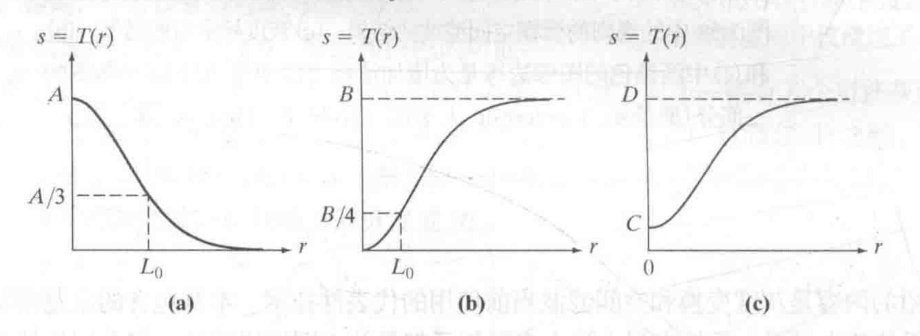
3.8 节的主要参考文献是由 L. A. Zadeh 撰写的关于模糊逻辑的三篇论文(Zadeh[1965, 1973, 1976])。如它们建立的模糊逻辑及其应用基础那样，这些论文是精心撰写的，有很高的阅读价值。关于模糊逻辑在图像处理中广泛应用的综述，见 Kerre and Nachtegae[2000]。3.8.4 节的中例子是以 Tizhuoosh[2000]描述的类似应用为基础的。3.8.5 节中的例子基本上来自 Russo and Ramponi[1994]。模糊集合应用于灰度变换和图像滤波的其他例子请分别参阅 Patrascu[2004]和 Nie and Barner[2006]。前面所述的从 1965 年到 2006 年的参考文献，是详细研究在图像处理中使用模糊集合的许多方法的起点。本章中讨论的许多方法的软件实现见 Gonzalez, Woods and Eddins[2004]。

标有星号的习题的详细解答可在本书的网站上找到。该网站还包含有基于本章内容的建议项目。

习题

★**3.1** 为了展开一幅图像的灰度，使其最低灰度为 C、最高灰度为 $L-1$，试给出一个单调的变换函数。

3.2 α 为正常数的指数式 $e^{-\alpha r^2}$ 对于构造平滑的灰度变换函数非常有用。从这个基本函数出发，构造具有下列图形形状的变换函数。所示的常数是输入参数，且你提出的变换必须以特定形式包含这些参数(为简化你的答案，第三条曲线中不需要参数 L_0)。



3.3 ★(a) 试给出实现图 3.2(a)所示的对比度拉伸变换的一个连续函数。此函数不仅包含参数 m，还必须包含参数 E，以便控制灰度值由低向高变化时函数的斜率。该函数应归一化，以使其最小值和最大值分别为 0 和 1。

(b) 画出作为参数 E 的函数的一组变换，固定值 $m = L/4$，其中 L 是图像中灰度的级数。

(c) 为使函数如图 3.2(b)的函数那样有效地执行，E 的最小值是什么？换句话说，你的函数与图 3.2(b)可以不同。它仅有产生一幅二值图像的相同结果。假设使用 8 比特图像进行处理，并使 $m = 128$。另外，令 C 是你正使用的计算机中所能表示的最小正数。

3.4 提出一组能够产生 4 比特单色图像所有单独比特平面的灰度分层变换(例如，变换函数 $T(r)$，当 r 在[0, 7]区间时，$T(r)=0$，而当 r 在[8, 15]区间时，$T(r)=15$，此函数可产生一幅 8 比特图像的第 4 比特平面图像)。

3.5 ★(a) 通常，如果将低阶比特平面的一半设为零值，对一幅图像的直方图有何影响?

(b) 如果将高阶比特平面的一半设为零值，对一幅图像的直方图有何影响?

★**3.6** 试解释为什么离散直方图均衡技术一般不能得到平坦的直方图。

3.7 假设对一幅数字图像进行直方图均衡处理。试证明(对直方图均衡后的图像)进行第二次直方图均衡处理的结果，与第一次直方图均衡处理的结果相同。

214 ~ 216

3.8 在某些应用中，将输入图像的直方图模型化为高斯概率密度函数很有用，高斯概率密度函数的形式为

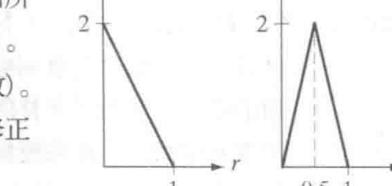
$$p_r(r)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(r-m)^2}{2\sigma^2}}$$

式中 m 和 σ 分别是高斯概率密度函数的均值和标准差。具体处理方法是将 m 和 σ 视为给定图像的平均灰度级和对比度。对于直方图均衡，你所用的变换函数是什么？

★3.9 假设图像的像素值连续，用一个例子说明有可能存在这样的情况，即由式(3.3-4)给出的变换函数满足 3.3.1 节中的条件(a)和(b)，但其反变换却不能满足条件(a)。

3.10 (a) 证明式(3.3.8)中给出的离散变换函数对直方图均衡处理满足 3.3.1 节中的条件(a)和(b)。

★(b) 证明仅在灰度 $r_k, k=0,1,\cdots,L-1$ 不丢失时，式(3.3-9)表示的离散直方图反变换才满足 3.3.1 节中的条件(a′)和(b)。



3.11 一幅灰度范围在[0, 1]内的图像的概率密度函数 $p_r(r)$ 如右图所示。现对此图像进行灰度变换，使其灰度分布为所示的 $p_z(z)$。假设灰度值连续，求实现这一要求的变换(表示为 r 和 z 的函数)。

★3.12 试提出一种如3.3.3 节中讨论的局部增强技术的局部直方图修正方法。

3.13 有两幅图像 $f(x,y)$ 和 $g(x,y)$，它们的直方图分别为 h_f 和 h_g。给出根据 h_f 和 h_g 确定如下直方图的条件，并解释如何获得每种情形下的直方图：

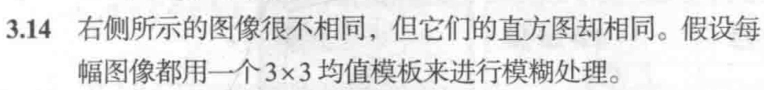
★(a) $f(x,y)+g(x,y)$。

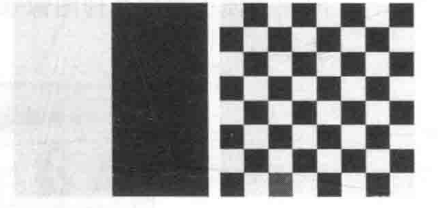
(b) $f(x,y)-g(x,y)$。

(c) $f(x,y)\times g(x,y)$。

(d) $f(x,y)\div g(x,y)$。

(e) $f(x,y)*g(x,y)$。





3.14 右侧所示的图像很不相同，但它们的直方图却相同。假设每幅图像都用一个 3×3 均值模板来进行模糊处理。

(a) 模糊后图像的直方图还相同吗？试解释原因。

(b) 如果你的答案是不相同，画出两个直方图。

3.15 线性空间滤波的实现要求在整幅图像中移动模板的中心点，并在每个位置计算模板系数与该位置相应像素值的乘积之和(见3.4节)。一个低通滤波器能用所有系数都为 1 的模板来实现，可使用所谓的盒状滤波器或移动均值算法，该方法由仅更新从一个位置到下一个位置变化的计算部分组成。

★(a) 以公式形式给出一个 $n\times n$ 滤波器的算法，说明涉及的计算规律，以及用于围绕图像移动模板的扫描序列。

(b) 大规模执行的计算量与盒状滤波器算法执行的计算量的比值，称为计算优势。在本例情况下求出计算优势，并画出它与 n 的关系曲线，其中 $n>1$。由于这两种处理方法的标定参数都是 $1/n^2$，因此在获得计算优势时根本不用考虑它的影响。假设图像有为 0 的外部边界，这个边界宽到足以让我们在分析中忽略边界的影响。

3.16 ★(a) 假如采用如式(3.4-2)定义的卷积方法用空间滤波模板 $w(x,y)$ 对图像 $f(x,y)$ 滤波，其中模板在两个空间方向上都小于图像。证明如下重要特性：如果模板的系数之和为 0，则得到的卷积阵列(被滤波的图像)中，所有元素之和也为 0(可以忽略计算的不准确性)。另外，你可以假设图像的边界已用合适数量的 0 填充。

(b) 如果用式(3.4-1)定义的相关方法来实现滤波，问得到的结果与(a)的结果相同吗？

216
~
217

3.17 讨论用一个3×3低通空间滤波器反复对一幅数字图像进行处理的结果，可以不考虑边界的影响。应用5×5滤波器时有何不同?

3.18 ★(a) 在3.5.2节中谈到，(相对于背景而言)孤立的暗或亮像素团块，在它们的面积小于中值滤波器面积的一半时，可被中值滤波器滤除(强迫为邻域的中值)。假设滤波器尺寸为$n \times n$，n为奇数，解释为什么会这样。

(b) 考虑一幅有不同像素团块的图像。假设一个团块中的所有点都要比背景亮或暗(同一团块中不同时存在这两种情形)，并且每个团块的面积不大于$n^2/2$。请问n符合什么条件时，有一个或多个这样的团块像(a)中所说的那样被分离出来?

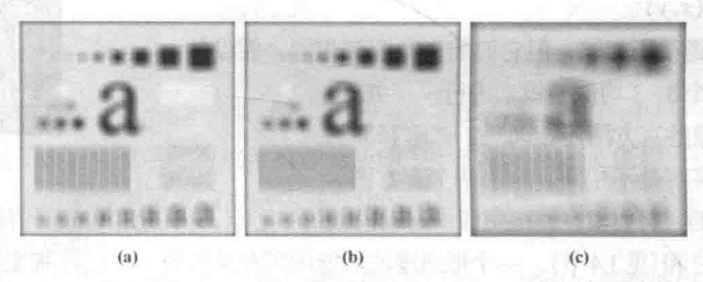
★3.19 (a) 试给出求一个$n \times n$邻域的中值的步骤。

(b) 试给出一种逐像素移动邻域的中心来更新中值的技术。

3.20 (a) 在字符识别应用中，文本页用图3.2(b)所示的阈值变换函数简化为二值图像。这遵循如下过程，即细化字符直到它们成为全"0"背景上的一串"1"。由于有噪声存在，故二值化和细化处理导致了字符的断开，断开缝隙宽1~3个像素。修复缝隙的一种方法是，对二值图像使用均值模板来模糊，进而在缝隙间连接非零像素。试求出能执行该任务的均值模板的最小尺寸。

(b) 连接缝隙后，为了转换回二值形式，要对图像进行阈值处理。根据你在(a)中得出的答案，完成这一任务且不再产生断线所要求的最小阈值是多少?

★3.21 以下三幅图像是分别经尺寸为$n = 23, 25$和45的方形均值模板处理后的模糊图像。图(a)和图(c)中左下角的垂直竖条被模糊，但竖条与竖条之间的分割仍然很清楚。然而，尽管产生这幅图像的模板要比处理图像(c)的小得多，但图(b)中的竖条却已经融在一起。试解释这一现象的原因。



(a) (b) (c)

3.22 考虑图3.34所示的应用，该应用的目的是消除图像中比$q \times q$像素方形包围的物体小的目标。假设想要将目标的平均灰度减少为原平均灰度的1/10。使用这种方法，那些目标可以接近背景灰度并用阈值法消除。给出平均模板的最小尺寸(奇数)，该模板仅对整幅图像处理一次就可将平均灰度级减少到所希望的程度。

3.23 在给定应用中，先对输入图像使用一个均值模板来减少噪声，然后再用一个拉普拉斯模板来增强图像中的细节。如果交换一下这两个步骤的顺序，结果是否会相同?

★3.24 证明如式(3.6-3)所示的拉普拉斯变换是各向同性的(旋转不变的)。你将需要下列轴旋转θ角的坐标方程：

$$x = x'\cos\theta - y'\sin\theta$$
$$y = x'\sin\theta + y'\cos\theta$$

式中(x, y)为未旋转的坐标，而(x', y')为旋转后的坐标。

217~219

★3.25 你在图3.38中看到中心为−8的拉普拉斯模板所得到的结果，要比中心为−4的模板所得到的结果清晰一些。详细说明其原因。

3.26 参考习题 3.25。

(a) 若使用一个类似于拉普拉斯的较大模板，如中心系数为–24 的 5×5 模板，问是否可得到更加清晰的结果？详细解释一下。

(b) 当该模板的大小与图像大小相等时，会发生什么情况？

3.27 给定一个 5×5 模板，使用该模板来对一幅图像执行非锐化模板处理。假设均值图像使用高斯滤波器得到。

★3.28 使用式(3.6-6)给出的拉普拉斯定义，证明从一幅图像中减去相应的拉普拉斯图像，等同于对图像进行非锐化模板处理。

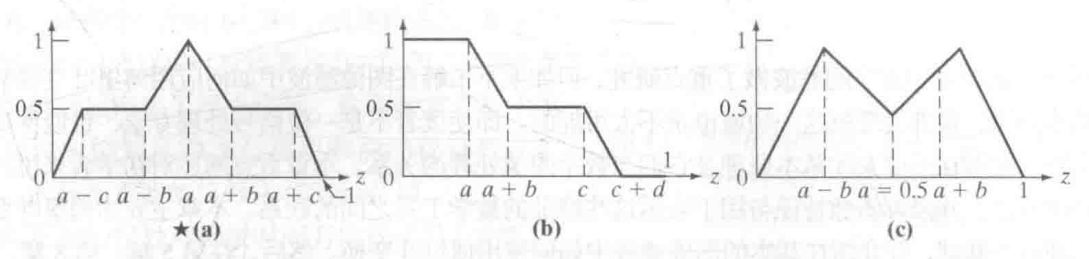
3.29 (a) 证明式(3.6-11)给出的梯度幅值是一种各向同性的操作(见习题 3.24)。

(b) 如果使用式(3.6-12)计算梯度，证明其将失去各向同性这一性质。

3.30 使用一台 CCD 电视摄像机对某一区域进行天候观测时，每 5 分钟拍摄一次数字图像并传送到中心场所。场景的照明白天为自然光，晚上为人造光，没有无照明的时间，因此总可以获得一幅图像。由于照明的范围总是在摄像机的线性工作范围，因此摄像机本身并不需要使用任何补偿装置。另外，使用数字技术对图像进行后处理并归一化，就使得图像与恒定照明时的图像等效。对此，请设计一种方法。你可以自由使用你希望的任何方法，但必须在设计中明确给出所有的假设。

3.31 证明图 3.46(d)中的交叉点由 $b=(a+c)/2$ 给出。

3.32 使用 3.8.2 节定义的模糊集合和图 3.46 中的基本隶属度函数，形成下列隶属度函数。



★3.33 在 3.8.5 节讨论的模糊滤波方法中，增大邻域的尺寸会产生什么影响？对你的回答解释原因(你可以用一个例子来支持你的答案)。

3.34 在灰度区间为[0, L–1]的带噪图像中，为降低脉冲噪声的影响，请如 3.8.5 节那样，设计一个基于规则的模糊系统。为简化问题，仅在一个 3×3 邻域中使用差值 d_2, d_4, d_6 和 d_8。在图像的任何位置，令 z_5 表示邻域中心的灰度。相应的输出灰度值应是 $z_5' = z_5 + v$，其中 v 是你的模糊系统的输出。也就是说，你的模糊系统的输出是一个用于减少可能出现在该 3×3 邻域中心的噪声峰值效应的校正因子。假设出现的噪声峰值间距大到无须关心在同一邻域内出现多个噪声峰值。这些峰值可以是暗的或亮的。全部使用三角形隶属度函数。

★(a) 给出该问题的一条模糊语句。

★(b) 指定 IF-THEN 规则和 ELSE 规则。

(c) 如图 3.57 那样，以图形方式指定隶属度函数。

(d) 如图 3.58 那样，给出规则集的图形表示。

(e) 给出一个类似于图 3.52 的模糊系统的简图。

219 ~ 220

第 4 章　频率域滤波

Filter: A device or material for suppressing or minimizing waves or oscillations of certain frequencies.

Frequency: The number of times that a periodic function repeats the same sequence of values during a unit variation of the independent variable.

Webster's New Collegiate Dictionary

引言

尽管在前一章中对空间滤波做了重点研究，但如果不了解在图像滤波中如何应用傅里叶变换和频率域的基本知识，要彻底理解这一领域也是不太可能的。即使读者不是一位信号处理专家，也能深刻理解这些主题，关键在于要关注基本原理及它们与数字图像处理的关系。本章重点阐述对初学者来说较为困难的表示方法，并强调图像特征与用于表示这些特征的数学工具之间的联系。本章主要为傅里叶变换的基本原理打下基础，并介绍在基本的图像滤波中如何使用傅里叶变换。然后，在第 5 章、第 8 章、第 10 章和第 11 章中，我们将讨论傅里叶变换的其他应用。本章的讨论从傅里叶变换的起源及其应用于数学、科学及工程的众多分支的要点开始。然后，我们从函数取样的基本原理开始，逐步推导一维和二维离散傅里叶变换及频率域处理的要领。在这段阐述中，我们还要接触几个重要的取样问题，例如，混淆，这些处理要求读者对频率域有所了解，且在本章会多次涉及。接下来是频率域滤波的公式表示，以及与第 3 章中讨论的空间平滑和锐化滤波技术相对应的内容。最后讨论图像处理中实现傅里叶变换的有关问题。因为 4.2 节到 4.4 节的内容是基本背景知识，熟悉一维信号处理，包括傅里叶变换、取样、混淆和卷积定理的读者，可直接跳到 4.5 节对二维傅里叶变换及其在数字图像处理中的应用的讨论。

4.1　背景

4.1.1　傅里叶级数和变换简史

法国数学家吉恩·巴普提斯特·约瑟夫·傅里叶(Jean Baptiste Joseph Fourier)于 1768 年生于巴黎与第戎之间的奥克塞里(Auxerre)镇。他被世人铭记的最大贡献记载在他于 1807 年发表的传记和 1822 年出版的 *La Théorie Analitique de la Chaleur*(热分析理论)一书中。此书由 Freeman(参见 Freeman [1878])在 55 年后翻译为英文。傅里叶在这个领域的贡献是，他指出任何周期函数都可以表示为不同频率的正弦和/或余弦之和的形式，每个正弦项和/或余弦项都乘以不同的系数(现在称该和为傅里叶级数)。无论函数多么复杂，只要它是周期的，并且满足某些适度的数学条件，都可以用这样的和来表示。

我们现在认为这是理所当然的，但在当时，这个概念首次提出后，复杂函数表示为简单正弦和余弦之和的概念很不直观(见图4.1)，所以傅里叶思想遭到质疑不足为奇。

甚至非周期函数(但该曲线下的面积是有限的)也可用正弦和/或余弦乘以加权函数的积分来表示。在这种情况下的公式就是傅里叶变换，其作用在多数理论和应用学科中甚至远大于傅里叶级数。用傅里叶级数或变换表示的函数特征完全可以通过傅里叶反变换来重建，且不会丢失任何信息。这是这种表示方法的最重要特征之一，因为它可以使我们工作于“傅里叶域”，而且在返回到函数的原始域时不会丢失任何信息。总之，傅里叶级数和变换是解决实际问题的工具，它作为基础工具被人们广泛地学习和利用。

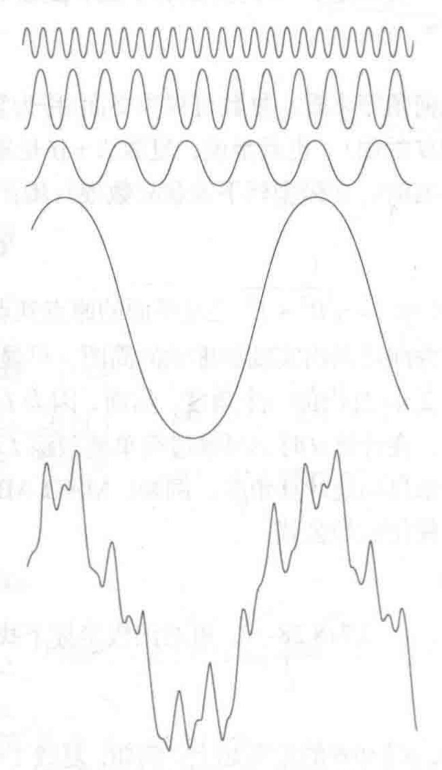


图 4.1　底部的函数是其上面 4 个函数的和。1807 年，傅里叶关于周期函数可以表示为正弦和余弦的加权和的概念遭到了人们的质疑

傅里叶概念的最初应用是在热扩散领域，在该领域，人们考虑用微分方程来表示热流，并使用这种方法首次得到了结论。在过去的一个世纪，尤其是后 50 年，傅里叶的思想使整个工业和学术界都空前繁荣。早在 20 世纪 60 年代(或更晚些)，数字计算的出现和快速傅里叶变换算法(FFT)的“发现”在信号处理领域产生了巨大变革。这两种核心技术第一次允许人们对从医学监视器和扫描仪到现代电子通信的异常重要的信号进行实际处理。

由于我们仅处理持续时间有限的函数(图像)，所以傅里叶变换是我们感兴趣的工具。下一节将介绍傅里叶变换和频率域。业已证明，傅里叶技术为研究和实现主要的图像处理算法提供了有意义且实用的方法。在某些情况下，这些方法与我们在第 3 章中介绍的方法类似。

4.1.2　关于本章中的例子

如第 3 章那样，本章中的多数图像滤波例子处理的是图像增强问题。例如，平滑和锐化与对比度处理技术一样，与图像增强有传统的联系。自然地，数字图像处理的初学者会发现图像增强很有趣并很容易理解。因此，在本章中以图像增强为例不仅可节省本书额外的章节，而且能为初学者介绍频率域滤波技术提供一个有效的工具。在第 5 章、第 8 章、第 10 章和第 11 章中，我们将对其他应用使用频率域处理方法。

4.2　基本概念

为简化本章的概念出现的进程，我们暂停介绍为后续章节奠定基础的一系列基本概念。

4.2.1　复数

复数 C 的定义如下：

$$C = R + \mathrm{j}I \tag{4.2-1}$$

222~224

式中，R 和 I 是实数，j 是一个等于−1 的平方根的虚数，即 $\mathrm{j}=\sqrt{-1}$。R 表示复数的实部，I 表示复数的虚部。实数是 $I=0$ 的复数的子集。复数 C 的共轭表示为 C^*，其定义是

$$C^* = R - \mathrm{j}I \tag{4.2-2}$$

从几何角度来看，复数可视为平面(称为复平面)上的一个点，其横坐标是实轴(R 的值)，其纵坐标是虚轴(I 的值)。也就是说，复数 $R+\mathrm{j}I$ 是复平面直角坐标系统中的点 (R,I)。

有时，在极坐标下表示复数很有用：

$$C = |C|(\cos\theta + \mathrm{j}\sin\theta) \tag{4.2-3}$$

式中，$|C|=\sqrt{R^2+I^2}$ 是复平面的原点到点 (R,I) 的向量的长度，θ 是该向量与实轴的夹角。在第一象限中为向量画出实轴和虚轴的简图，可显示出 $\tan\theta=(I/R)$ 或 $\theta=\arctan(I/R)$。arctan 函数返回区间 $[-\pi/2,\pi/2]$ 内的一个角度。然而，因为 I 和 R 可独立为正和为负，因此我们需要得到全域 $[-\pi,\pi]$ 内的角度。在计算 θ 时，可通过简单地跟踪 I 和 R 的符号来完成。许多编程语言可通过调用四象限反正切函数来自动地计算角度。例如，MATLAB 为此目的提供了函数 `atan2(Imag,Real)`。

使用欧拉公式

$$\mathrm{e}^{\mathrm{j}\theta} = \cos\theta + \mathrm{j}\sin\theta \tag{4.2-4}$$

式中，$\mathrm{e}=2.718\,28\cdots$，可给出极坐标下我们很熟悉的复数表示：

$$C = |C|\mathrm{e}^{\mathrm{j}\theta} \tag{4.2-5}$$

式中，$|C|$ 和 θ 的定义如上。例如，复数 1 + j2 的极坐标表示是 $\sqrt{5}\mathrm{e}^{\mathrm{j}\theta}$，其中 $\theta=64.4°$ 或 1.1 弧度(rad)。前面的公式还可以用于复函数。例如，变量 u 的复函数 $F(u)$ 可表示为 $F(u)=R(u)+\mathrm{j}I(u)$，其中 $R(u)$ 和 $I(u)$ 分别是实分量函数和虚分量函数。如前所述，复共轭函数是 $F^*(u)=R(u)-\mathrm{j}I(u)$，幅值是 $|F(u)|=\sqrt{R(u)^2+I(u)^2}$，角度是 $\theta(u)=\arctan[I(u)/R(u)]$。在本章和下一章中，我们还会多次讨论复函数。

4.2.2 傅里叶级数

如 4.1.1 节指出的那样，具有周期 T 的连续变量 t 的周期函数 $f(t)$ 可描述为乘以适当系数的正弦和余弦之和。我们知道，这个和就是傅里叶级数，它具有如下形式：

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{j}\frac{2\pi n}{T}t} \tag{4.2-6}$$

式中，

$$c_n = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\mathrm{e}^{-\mathrm{j}\frac{2\pi n}{T}t}\,\mathrm{d}t, \quad n=0,\ \pm 1,\ \pm 2,\ \cdots \tag{4.2-7}$$

是系数。式(4.2-6)可展开为正弦与余弦之和这一事实来自欧拉公式(4.2-4)。在本章的后面，我们将回到傅里叶级数。

4.2.3 冲激及其取样特性

> 冲激并不是通常意义上的函数，其更准确的名称是分布或广义函数。但是，我们在文献中通常会看到冲激函数、δ 函数和狄拉克 δ 函数这样的名称，尽管它们都名不副实。

线性系统和傅里叶变换研究的核心是冲激及其取样特性。连续变量 t 在 $t=0$ 处的单位冲激表示为 $\delta(t)$，其定义是

224
~
225

$$\delta(t)=\begin{cases}\infty, & t=0\\ 0, & t\neq 0\end{cases} \tag{4.2-8a}$$

它还被限制为满足等式

$$\int_{-\infty}^{\infty}\delta(t)\mathrm{d}t=1 \tag{4.2-8b}$$

物理上，如果我们把 t 解释为时间，那么一个冲激可视为幅度无限、持续时间为 0、具有单位面积的尖峰信号。一个冲激具有关于如下积分的所谓取样（sifting）特性：

$$\int_{-\infty}^{\infty}f(t)\delta(t)\mathrm{d}t=f(0) \tag{4.2-9}$$

假设 $f(t)$ 在 $t=0$ 处是连续的，这是实际中满足的一个典型条件。取样特性得到函数 $f(t)$ 在冲激位置（例如，在前面的公式中为原点 $t=0$）的值。取样特性的一种更为一般的说明涉及位于任意点 t_0 的冲激，表示为 $\delta(t-t_0)$。在这种情况下，取样特性变为

> sift 字面上的意思是分隔，或通过一个筛子分拣。

$$\int_{-\infty}^{\infty}f(t)\delta(t-t_0)\mathrm{d}t=f(t_0) \tag{4.2-10}$$

它在冲激位置 t_0 处得到一个函数值。例如，如果 $f(t)=\cos(t)$，使用式(4.2-10)中的冲激 $\delta(t-\pi)$ 得到结果 $f(\pi)=\cos(\pi)=-1$。取样概念的作用立即变得很明显。

令 x 表示一个离散变量。单位离散冲激 $\delta(x)$ 在离散系统中的作用与处理连续变量时冲激 $\delta(t)$ 的作用相同。其定义如下：

$$\delta(x)=\begin{cases}1, & x=0\\ 0, & x\neq 0\end{cases} \tag{4.2-11a}$$

很明显，该定义也满足式(4.2-8b)的离散等效形式：

$$\sum_{x=-\infty}^{\infty}\delta(x)=1 \tag{4.2-11b}$$

离散变量的取样特性有如下形式：

$$\sum_{x=-\infty}^{\infty}f(x)\delta(x)=f(0) \tag{4.2-12}$$

或者，更一般地用位置 $x=x_0$ 处的离散冲激，

$$\sum_{x=-\infty}^{\infty}f(x)\delta(x-x_0)=f(x_0) \tag{4.2-13}$$

如前面一样，我们看到，取样特性可简单地得到冲激位置处的函数值。图4.2以图解方式显示了单位离散冲激。与其连续形式不同的是，离散冲激是一个普通函数。

本节后面特别感兴趣的是冲激串 $s_{\Delta T}(t)$，它定义为无限多个离散的周期冲激单元 ΔT 之和：

$$s_{\Delta T}(t)=\sum_{n=-\infty}^{\infty}\delta(t-n\Delta T) \tag{4.2-14}$$

图 4.3 显示了一个冲激串。冲激可以是连续的或离散的。

225~226

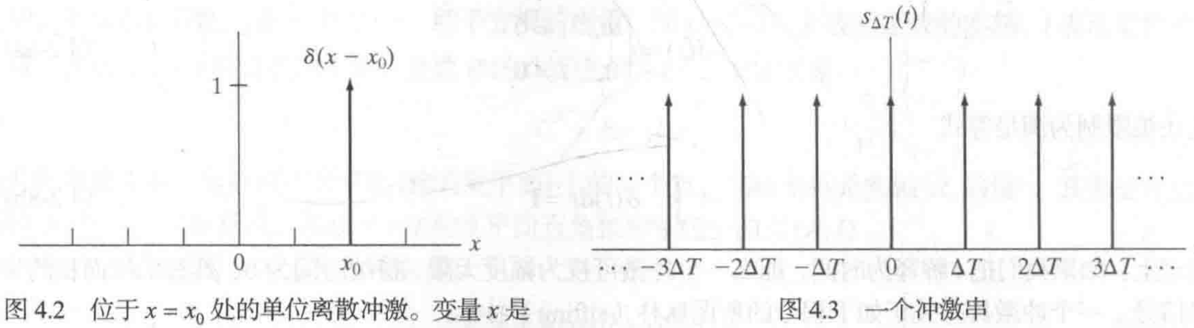


图 4.2 位于 $x = x_0$ 处的单位离散冲激。变量 x 是离散的，任何不在 $x = x_0$ 处的 δ 都是 0

图 4.3 一个冲激串

4.2.4 连续变量函数的傅里叶变换

由 $\Im\{f(t)\}$ 表示的连续变量 t 的连续函数 $f(t)$ 的傅里叶变换由下式定义[①]：

$$\Im\{f(t)\} = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t \tag{4.2-15}$$

式中，μ 也是一个连续变量。因为 t 被积分过，故 $\Im\{f(t)\}$ 仅是 μ 的函数。为了明确表示这一事实，我们把傅里叶变换写成 $\Im\{f(t)\} = F(\mu)$；也就是说，为方便起见，$f(t)$ 的傅里叶变换可写为

$$F(\mu) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t \tag{4.2-16}$$

相反，给定 $F(\mu)$，通过傅里叶反变换可以得到 $f(t)$，即 $f(t) = \Im^{-1}\{F(\mu)\}$，写为

$$f(t) = \int_{-\infty}^{\infty} F(\mu)\mathrm{e}^{\mathrm{j}2\pi\mu t}\mathrm{d}\mu \tag{4.2-17}$$

这里，我们使用了这样一个事实，即在反变换中变量 μ 被积分过，并简单地写成 $f(t)$，而不是更麻烦地表示为 $f(t) = \Im^{-1}\{F(\mu)\}$。式(4.2-16)和式(4.2-17)共同称为**傅里叶变换对**。它们指出了在 4.1 节中提到的一个重要事实，即一个函数可以由其变换来恢复。

使用欧拉公式，我们可以把式(4.2-16)表示为

$$F(\mu) = \int_{-\infty}^{\infty} f(t)\left[\cos(2\pi\mu t) - \mathrm{j}\sin(2\pi\mu t)\right]\mathrm{d}t \tag{4.2-18}$$

我们看到，如果 $f(t)$ 是实数，那么其变换通常是复数。注意，傅里叶变换是 $f(t)$ 乘以正弦项的展开，正弦项的频率由 μ 的值决定(如早些时候提到过的，变量 t 被积分过)。因为积分后左边剩下的唯一变量是频率，故我们说傅里叶变换域是频率域。在本章稍后，我们将详细讨论频率域及其特性。在我们的讨论中，t 可以表示任何连续变量，频率变量 μ 的单位取决于 t 的单位。例如，如果 t 表示单位为秒的时间，则 μ 的单位为周/秒，或赫兹(Hz)。如果 t 表示的是以米为单位的距离，则 μ 的单位是周/米，等等。换句话说，频率域的单位与输入变量的单位周期无关。

> 为与前两章中使用的术语保持一致，并在本章后面用于图像，当变量 t 用于说明图像时，我们一般将变量 t 的域称为空间域。



① 声明傅里叶变换存在的条件一般来说很复杂(Champeney[1987])，但其存在的充分条件是 $f(t)$ 的绝对值或 $f(t)$ 的平方的积分是有限的。实践中傅里叶变换的存在性通常不是问题，但理想信号除外，因为这种正弦信号会无限扩展。这些内容可使用通用的冲激函数来处理。我们的主要兴趣在于离散傅里叶变换对，它是所有有限函数存在的保证，后面我们很快会看到这一点。

例 4.1　求一个简单函数的傅里叶变换。

图4.4(a)中的函数的傅里叶变换可由式(4.2-16)得出：

$$
\begin{aligned}
F(\mu) &= \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-j2\pi\mu t}\mathrm{d}t = \int_{-W/2}^{W/2} A\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t \\
&= \frac{-A}{\mathrm{j}2\pi\mu}\left[\mathrm{e}^{-\mathrm{j}2\pi\mu t}\right]_{-W/2}^{W/2} = \frac{-A}{\mathrm{j}2\pi\mu}\left[\mathrm{e}^{-\mathrm{j}2\pi\mu W} - \mathrm{e}^{\mathrm{j}2\pi\mu W}\right] \\
&= \frac{A}{\mathrm{j}2\pi\mu}\left[\mathrm{e}^{\mathrm{j}\pi\mu W} - \mathrm{e}^{-\mathrm{j}\pi\mu W}\right] = AW\frac{\sin(\pi\mu W)}{(\pi\mu W)}
\end{aligned}
$$

式中，我们利用了三角恒等式 $\sin\theta = (\mathrm{e}^{\mathrm{j}\theta} - \mathrm{e}^{-\mathrm{j}\theta})/2\mathrm{j}$ 。在这种情况下，傅里叶变换的复数项精细地合并为一个实正弦函数。前面表达式的最后一步是熟知的 sinc 函数：

$$
\mathrm{sinc}(m) = \frac{\sin(\pi m)}{(\pi m)} \tag{4.2-19}
$$

式中，$\mathrm{sinc}(0) = 1$，对于 m 的所有其他整数值，$\mathrm{sinc}(m) = 0$。图4.4(b)显示了 $F(\mu)$ 的曲线。

通常，傅里叶变换包含复数项，且为显示目的，通常处理该变换的幅值(一个实量)，该幅值称为傅里叶谱或频谱：

$$
|F(\mu)| = AW\left|\frac{\sin(\pi\mu W)}{(\pi\mu W)}\right|
$$

图4.4(c)显示了作为频率的函数的 $|F(\mu)|$ 的曲线。我们注意的关键特性是 $F(\mu)$ 和 $|F(\mu)|$ 的零值位置与“盒状”函数的宽度 W 成反比，即波瓣的高度随函数距原点的距离降低，并且函数在 μ 值的正方向和负方向上无限扩展。如我们稍后将会看到的那样，这些性质对解释图像的二维傅里叶变换谱十分有用。

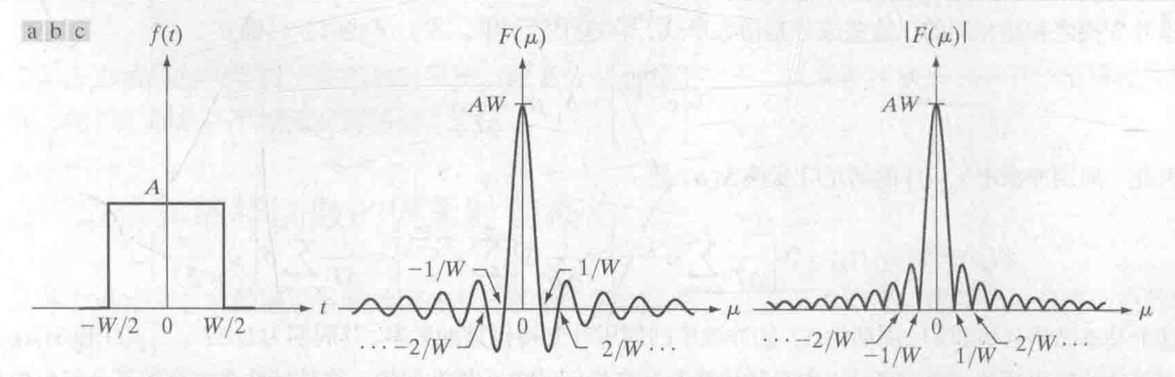


图 4.4　(a)一个简单的函数；(b)该函数的傅里叶变换；(c)该函数的谱。所有函数在两个方向上都无限扩展

例 4.2　冲激和冲激串的傅里叶变换。

位于原点的单位冲激的傅里叶变换由式(4.2-16)给出：

$$
F(\mu) = \int_{-\infty}^{\infty} \delta(t)\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{j}2\pi\mu t}\delta(t)\mathrm{d}t = \mathrm{e}^{-\mathrm{j}2\pi\mu 0} = \mathrm{e}^{0} = 1
$$

式中，第三步由式(4.2-9)中的取样特性得出。这样，我们看到，位于空间域原点的一个冲激的傅里叶变换，在频率域是一个常数。类似地，位于 $t = t_0$ 的一个冲激的傅里叶变换是

$$
F(\mu) = \int_{-\infty}^{\infty} \delta(t - t_0)\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{j}2\pi\mu t}\delta(t - t_0)\mathrm{d}t = \mathrm{e}^{-\mathrm{j}2\pi\mu t_0} = \cos(2\pi\mu t_0) - \mathrm{j}\sin(2\pi\mu t_0)
$$

上式的第三步由式(4.2-10)中的取样特性得到，最后一行由欧拉公式得到。最后两行是以复平面原点为中心的单位圆的等效表示。

228 ~ 230

在 4.3 节，我们将利用周期冲激串的傅里叶变换。得到这个变换并不像我们得到刚才显示的单个冲激的变换那样简单。然而，理解如何推导一个冲激串的变换十分重要，因此这里我们花一些时间来详细推导它。我们从仅在形式上与式(4.2-16)和(4.2-17)的指数符号上有些不同的注释开始。这样，如果函数 $f(t)$ 有傅里叶变换 $F(\mu)$，则在点 t 求值后函数是 $F(t)$，它一定有变换 $f(-\mu)$。使用这种对称特性和上面给出的冲激 $\delta(t-t_0)$ 的傅里叶变换 $e^{-j2\pi\mu t_0}$，可得函数 $e^{-j2\pi\mu t_0}$ 的变换为 $\delta(-\mu-t_0)$。令 $-t_0=a$，可得 $e^{-j2\pi at}$ 的变换是 $\delta(-\mu+a)=\delta(\mu-a)$，其中最后一步是正确的，因为 δ 仅在 $\mu=a$ 时不为零，对于 $\delta(-\mu+a)$ 或 $\delta(\mu-a)$，结果均相同，因此这两种形式是等价的。

式(4.2-14)中的冲激串 $s_{\Delta T}(t)$ 是周期为 ΔT 的周期函数，因此由 4.2.2 节可知它可表示为一个傅里叶级数：

$$s_{\Delta T}(t)=\sum_{n=-\infty}^{\infty}c_n e^{j\frac{2\pi n}{\Delta T}t}$$

式中，

$$c_n=\frac{1}{\Delta T}\int_{-\Delta T/2}^{\Delta T/2}s_{\Delta T}(t)e^{-j\frac{2\pi n}{\Delta T}t}\mathrm{d}t$$

参考图4.3，我们看到区间 $[-\Delta T/2,\Delta T/2]$ 的积分仅包含位于原点的冲激 $s_{\Delta T}(t)$。因此，上式变为

$$c_n=\frac{1}{\Delta T}\int_{-\Delta T/2}^{\Delta T/2}\delta(t)e^{-j\frac{2\pi n}{\Delta T}t}\mathrm{d}t=\frac{1}{\Delta T}e^0=\frac{1}{\Delta T}$$

然后，傅里叶级数可展开为

$$s_{\Delta T}(t)=\frac{1}{\Delta T}\sum_{n=-\infty}^{\infty}e^{j\frac{2\pi n}{\Delta T}t}$$

我们的目的是得到该表达式的傅里叶变换。因为求和是线性过程，和的傅里叶变换与求各个成分的傅里叶变换之和是相同的。这些成分是指数形式，在这个例子中，我们早些时候已确立

$$\Im\left\{e^{j\frac{2\pi n}{\Delta T}t}\right\}=\delta\left(\mu-\frac{n}{\Delta T}\right)$$

因此，周期冲激串 $s_{\Delta T}(t)$ 的傅里叶变换 $S(\mu)$ 是

$$S(\mu)=\Im\{s_{\Delta T}(t)\}=\Im\left\{\frac{1}{\Delta T}\sum_{n=-\infty}^{\infty}e^{j\frac{2\pi n}{\Delta T}t}\right\}=\frac{1}{\Delta T}\Im\left\{\sum_{n=-\infty}^{\infty}e^{j\frac{2\pi n}{\Delta T}t}\right\}=\frac{1}{\Delta T}\sum_{n=-\infty}^{\infty}\delta\left(\mu-\frac{n}{\Delta T}\right)$$

这个基本结果告诉我们，周期为 ΔT 的冲激串的傅里叶变换仍是冲激串，其周期为 $1/\Delta T$。$s_{\Delta T}(t)$ 和 $S(\mu)$ 之间周期的这种反比关系与图 4.4 中盒状函数及其变换间的关系是类似的。这种特性在本章的其余部分中扮演着主要角色。

4.2.5 卷积

在讨论之前，我们还需要建立一个模块。我们在 3.4.2 节介绍过卷积的概念。在那一节，我们已了解两个函数的卷积涉及一个函数关于原点翻转(旋转180°)并滑过另一个函数。在滑动过程中的每一个位移处，我们执行计算，在第 3 章的情况下是计算乘积之和。在当前讨论中，我们的兴趣在于具有连续变量 t 的两个连续函数 $f(t)$ 和 $h(t)$ 的卷积，因此我们必须用积分代替求和。像之前那样，这两个函数的卷积由算子★表示，卷积定义如下：

$$f(t)\bigstar h(t)=\int_{-\infty}^{\infty}f(\tau)h(t-\tau)\mathrm{d}\tau \tag{4.2-20}$$

230 ~ 231

式中，负号表示刚才提及的翻转，t 是一个函数滑过另一个函数的位移，而 τ 是积分假变量。现在我们假定函数从 $-\infty$ 扩展到 ∞ 。

在图 3.4.2 中我们说明了卷积的基本机制，且在本章稍后和第 5 章中我们会再次说明它。此时，我们的兴趣在于寻找式(4.2-20)的傅里叶变换。我们从式(4.2-15)开始：

$$\Im\{f(t)\bigstar h(t)\}=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}f(\tau)h(t-\tau)\mathrm{d}\tau\right]\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t=\int_{-\infty}^{\infty}f(\tau)\left[\int_{-\infty}^{\infty}h(t-\tau)\mathrm{e}^{-\mathrm{j}2\pi\mu t}\mathrm{d}t\right]\mathrm{d}\tau$$

方括号中的项是 $h(t-\tau)$ 的傅里叶变换。我们在本章稍后将证明 $\Im\{h(t-\tau)\}=H(\mu)\mathrm{e}^{-\mathrm{j}2\pi\mu\tau}$，其中 $H(\mu)$ 是 $h(t)$ 的傅里叶变换。在前一公式中利用这一事实可得

> 如果颠倒 $f(t)$ 和 $h(t)$ 的顺序，可以得到相同的结果，故卷积满足交换律。

$$\Im\{f(t)\bigstar h(t)\}=\int_{-\infty}^{\infty}f(\tau)\left[H(\mu)\mathrm{e}^{-\mathrm{j}2\pi\mu\tau}\right]\mathrm{d}\tau=H(\mu)\int_{-\infty}^{\infty}f(\tau)\mathrm{e}^{-\mathrm{j}2\pi\mu\tau}\mathrm{d}\tau=H(\mu)F(\mu)$$

回顾 4.2.4 节，我们将 t 所在的域称为空间域，而将 μ 所在的域称为频率域，前面的公式告诉我们，空间域中两个函数的卷积的傅里叶变换，等于两个函数的傅里叶变换在频率域中的乘积。反过来，如果有两个变换的乘积，那么我们可以通过计算傅里叶反变换得到空间域的卷积。换句话说，$f(t)\bigstar h(t)$ 和 $H(\mu)F(\mu)$ 是傅里叶变换对。这一结果是卷积定理的一半，并可写为

$$f(t)\bigstar h(t)\Leftrightarrow H(\mu)F(\mu)\tag{4.2-21}$$

双箭头用于指示右边的表达式是通过对左边的表达式执行傅里叶变换得到的，而左边的表达式是通过求右边的表达式的傅里叶反变换得到的。

遵循类似的推导，可得到卷积定理的另一半：

$$f(t)h(t)\Leftrightarrow H(\mu)\bigstar F(\mu)\tag{4.2-22}$$

它说明频率域的卷积类似于空间域的乘积，两者分别与傅里叶正、反变换相联系。如本章稍后提及的那样，卷积定理是频率域滤波的基础。

4.3　取样和取样函数的傅里叶变换

本节利用 4.2 节的概念系统地阐述取样的数学基础，首先介绍基本原理，然后介绍取样函数的傅里叶变换。

4.3.1　取样

用计算机处理之前，连续函数必须转换为离散值序列。如 2.4 节介绍的那样，这是用取样和量化来完成的。在下面的讨论中，我们将详细介绍取样。

参考图4.5，考虑一个连续函数 $f(t)$ ，我们希望以自变量 t 的均匀间隔(ΔT)取样。我们假定函数关于 t 从 $-\infty$ 扩展到 ∞。如 4.2.3 节讨论的那样，模拟取样的一种方法是，用一个 ΔT 单位间隔的冲激串作为取样函数去乘以 $f(t)$ ，即

$$\tilde{f}(t)=f(t)s_{\Delta T}(t)=\sum_{n=-\infty}^{\infty}f(t)\delta(t-n\Delta T)\tag{4.3-1}$$

式中， $\tilde{f}(t)$ 表示取样后的函数。这个和式的每个成分都是由该冲激位置处 $f(t)$的值加权后的冲激，

231 ~ 234

如图 4.5(c)所示。每个取样值由加权后的冲激“强度”给出，我们可通过积分得到它。也就是说，序列中的任意取样值 f_k 由下式给出：

$$f_k = \int_{-\infty}^{\infty} f(t)\delta(t-k\Delta T)\mathrm{d}t = f(k\Delta T) \tag{4.3-2}$$

按 ΔT 单位间隔取样表明取样率等于 $1/\Delta T$。如果 ΔT 的单位为秒，则采样率的单位是取样数/秒。如果 ΔT 的单位是米，则取样率的单位是取样数/米，以此类推。

式中，我们利用了式(4.2-10)中 δ 的取样特性。式(4.3-2)对任何整数 $k=\cdots,-2,-1,0,1,2,\cdots$ 都成立。图4.5(d)显示了结果，它由原始函数的等间隔取样组成。

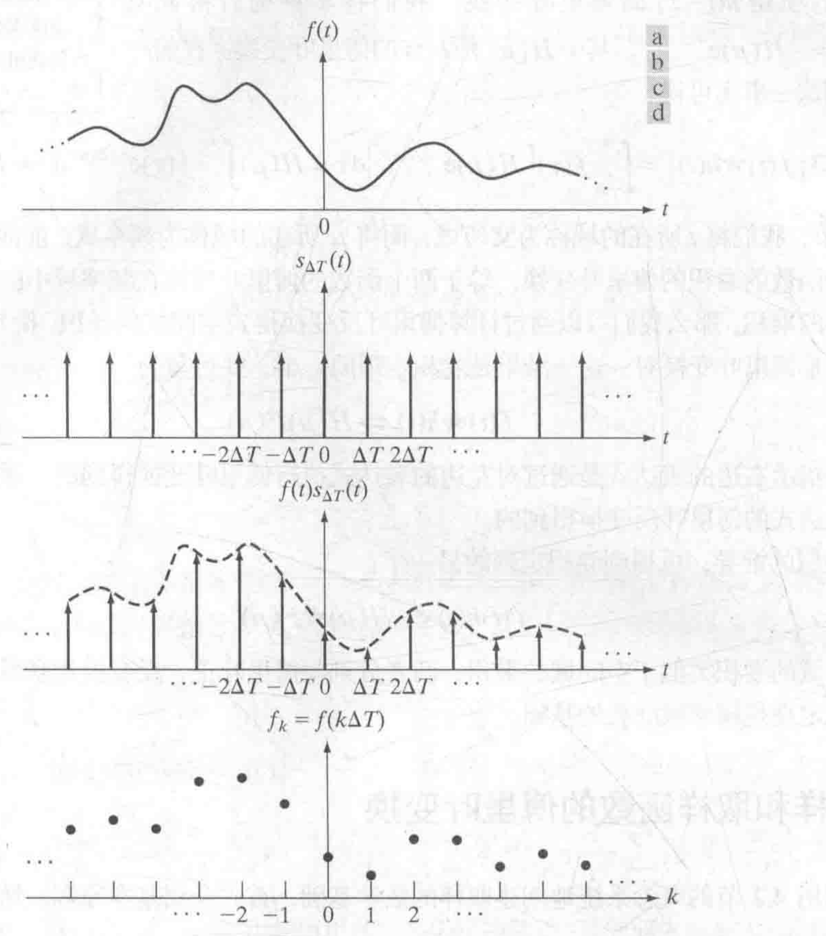


图 4.5 (a)一个连续函数；(b)用于模拟取样过程的冲激串；(c)图(a)和图(b)相乘形成的取样后的函数；(d)由积分并使用冲激的取样特性得到的取样值［图(c)中的虚线仅供参考，它不是数据的一部分］

4.3.2 取样函数的傅里叶变换

令 $F(\mu)$ 代表连续函数 $f(t)$ 的傅里叶变换。如前节讨论的那样，取样后的相应函数 $\tilde{f}(t)$ 是 $f(t)$ 与一个冲激串的乘积。由 4.2.5 节的卷积定理可知，空间域两个函数乘积的傅里叶变换是两个函数在频率域的卷积。这样，取样后的函数 $\tilde{f}(t)$ 的傅里叶变换 $\tilde{F}(\mu)$ 是

$$\tilde{F}(\mu) = \Im\{\tilde{f}(t)\} = \Im\{f(t)s_{\Delta T}(t)\} = F(\mu)\bigstar S(\mu) \tag{4.3-3}$$

其中，由例 4.2 可知

234

$$S(\mu) = \frac{1}{\Delta T}\sum_{n=-\infty}^{\infty}\delta\left(\mu - \frac{n}{\Delta T}\right) \tag{4.3-4}$$

是冲激串 $s_{\Delta T}(t)$ 的傅里叶变换。由式(4.2-20)中的定义，可直接得到 $F(\mu)$ 和 $S(\mu)$ 的卷积：

$$\begin{aligned}\tilde{F}(\mu) &= F(\mu)\bigstar S(\mu) = \int_{-\infty}^{\infty} F(\tau)S(\mu-\tau)\,\mathrm{d}\tau \\ &= \frac{1}{\Delta T}\int_{-\infty}^{\infty} F(\tau)\sum_{n=-\infty}^{\infty}\delta\left(\mu-\tau-\frac{n}{\Delta T}\right)\mathrm{d}\tau = \frac{1}{\Delta T}\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} F(\tau)\delta\left(\mu-\tau-\frac{n}{\Delta T}\right)\mathrm{d}\tau \\ &= \frac{1}{\Delta T}\sum_{n=-\infty}^{\infty} F\left(\mu-\frac{n}{\Delta T}\right)\end{aligned} \tag{4.3-5}$$

式中，最后一步来自于由式(4.2-10)给出的冲激取样特性。

式(4.3-5)中最后一行的求和表明，取样后的函数 $\tilde{f}(t)$ 的傅里叶变换 $\tilde{F}(\mu)$ 是 $F(\mu)$ 的一个副本的无限、周期序列，也是原始连续函数的傅里叶变换。副本间的间隔由 $1/\Delta T$ 的值决定。很明显，虽然 $\tilde{f}(t)$ 是取样后的函数，但其变换 $\tilde{F}(\mu)$ 是连续的，因为它由 $F(\mu)$ 的几个副本组成，所以 $F(\mu)$ 是一个连续函数。

图 4.6 是前面结果的图示总结[①]。图 4.6(a)是函数 $f(t)$ 的傅里叶变换 $F(\mu)$ 的简图，图 4.6(b)显示了取样后的函数的变换 $\tilde{F}(\mu)$。如前节提到的那样，$1/\Delta T$ 是用于生成取样后的函数的取样率。因此，在图 4.6(b)中，取样率要足够高，以便在各个周期之间提供有效的间隔，并保持 $F(\mu)$ 的完整性。在图 4.6(c)中，取样率刚好足以保持 $F(\mu)$，但在图 4.6(d)中，取样率低于保持不同 $F(\mu)$ 副本的最小取样率要求，因此不能保持原始变换。图4.6(b)是对信号过取样后的结果，图4.6(c)和(d)分别是对信号临界取样和欠取样后的结果。这些概念是下一节内容的基础。

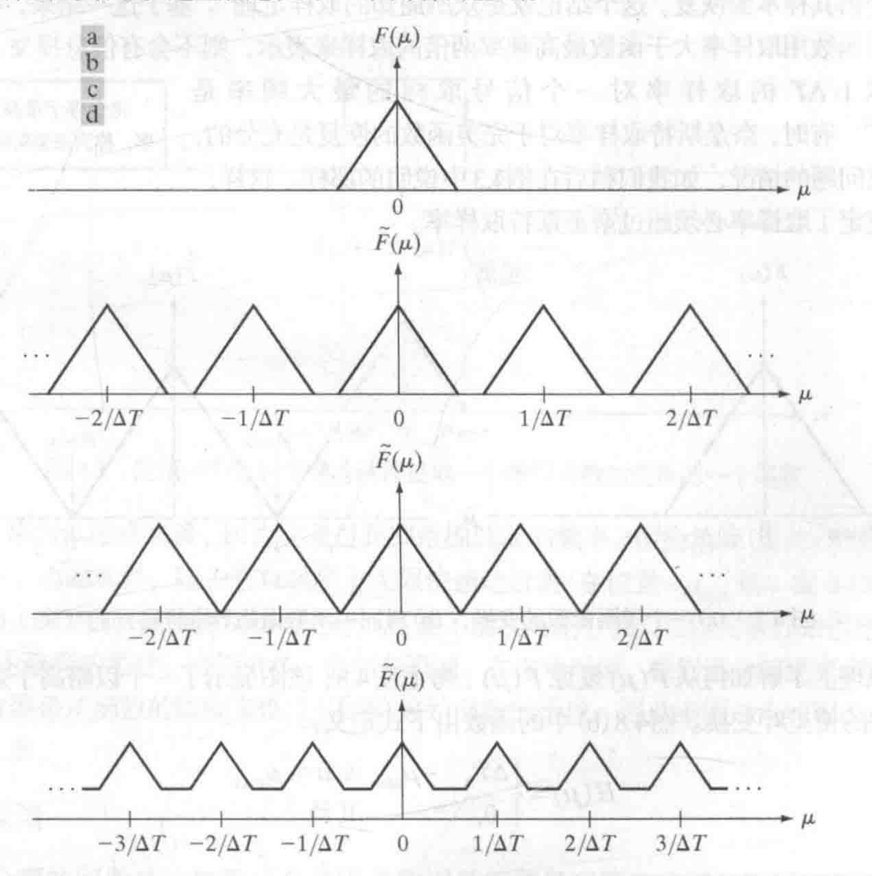


图 4.6　(a)一个带限函数的傅里叶变换；(b)~(d)分别在过取样、临界取样、欠取样条件下取样后的函数的傅里叶变换

235~236

① 为清楚起见，图 4.6 中傅里叶变换的草图及本章中的其他类似图形，忽略了变换一般是复函数这一事实。

4.3.3 取样定理

2.4 节简单地介绍了取样的概念。现在我们考虑形式上的取样过程，并明确一个连续函数完全可由其取样集合来恢复的条件。

对于以原点为中心的有限区间(带宽) $[-\mu_{max}, \mu_{max}]$ 之外的频率值，其傅里叶变换为零的函数 $f(t)$ 称为带限函数。图 4.7(a) 就是这样一个函数，它是图 4.6(a) 的一个放大部分。类似地，图 4.7(b) 是图 4.6(c) 所示临界取样后函数的傅里叶变换的详细视图。较低的 $1/\Delta T$ 值将导致 $\tilde{F}(\mu)$ 中的周期融合，较高的 $1/\Delta T$ 值在周期之间会提供更为清晰的间隔。

如果能从 $\tilde{F}(\mu)$ 中包含的这个函数的副本的周期序列中分离出 $F(\mu)$ 的一个副本，就可以从取样后的版本复原 $f(t)$，其中 $\tilde{F}(\mu)$ 是取样后的函数 $\tilde{f}(t)$ 的傅里叶变换。回顾前节的讨论，$\tilde{F}(\mu)$ 是周期为 $1/\Delta T$ 的连续周期函数。因此，我们需要一个完整的周期来表征整个变换。这意味着我们用傅里叶反变换从一个单周期就可恢复 $f(t)$。

若副本间的间距足够大(见图 4.6)，则有可能从 $\tilde{F}(\mu)$ 中提取单个周期使其等于 $F(\mu)$。根据图 4.7(b)，若 $1/2\Delta T > \mu_{max}$ 或

$$\frac{1}{\Delta T} > 2\mu_{max} \tag{4.3-6}$$

则可保证有足够大的间距。该式指出，如果以超过函数最高频率的两倍的取样率来获得样本，连续带限函数能完全由其样本集恢复。这个结论就是众所周知的取样定理[①]。基于这一结果，我们可以说，若一个连续带限函数用取样率大于函数最高频率两倍的取样来表示，则不会有信息损失。反之，我们也可以说，以 $1/\Delta T$ 的取样率对一个信号取样的最大频率是 $\mu_{max} = 1/2\Delta T$。有时，奈奎斯特取样率对于完美函数的恢复是充分的，但也存在导致问题的情况，如我们稍后在例4.3中说明的那样。这样，取样定理就规定了取样率必须超过奈奎斯特取样率。

> 完全等于最高频率的两倍的取样率，称为奈奎斯特取样率。

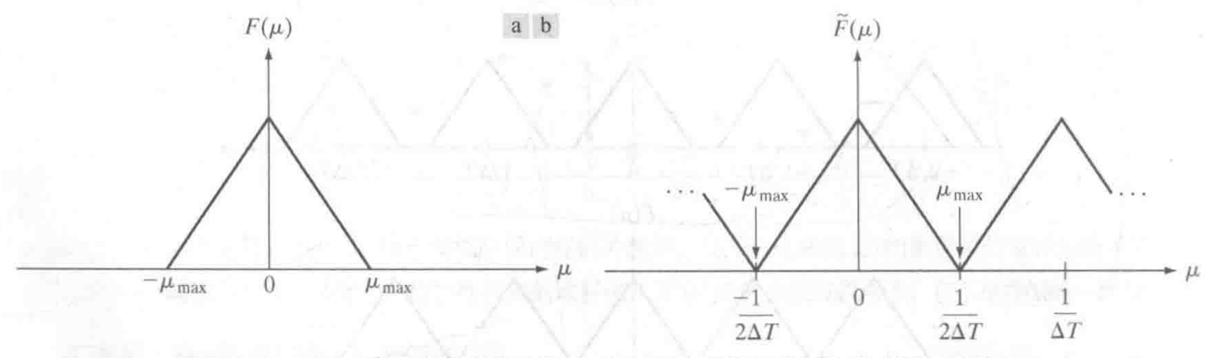


图 4.7 (a)一个带限函数的变换，(b)对同一函数做临界取样得到的变换

为了从原理上了解如何从 $\tilde{F}(\mu)$ 复原 $F(\mu)$，考虑图4.8，该图显示了一个以略高于奈奎斯特取样率取样后的函数的傅里叶变换。图4.8(b) 中的函数由下式定义：

$$H(\mu) = \begin{cases} \Delta T, & -\mu_{max} \leqslant \mu \leqslant \mu_{max} \\ 0, & \text{其他} \end{cases} \tag{4.3-7}$$

236 ~ 238

① 取样定理是数字信号处理理论的基石。该定理由 Harry Nyquist(哈里·奈奎斯特，美国贝尔实验室的一位科学家和工程师)于 1928 年最先提出。同样来自贝尔实验室的 Claude E. Shannon(克劳德·E·山农)于 1949 年正式地证明了该定理。早期数字计算系统和现代通信的出现，使得人们开始研究处理数字(取样后的)数据的方法，从而导致了人们于 20 世纪 40 年代后期重新开始关注取样定理。

当乘以图 4.8(a) 中的周期序列时，该函数就分隔了以原点为中心的周期。然后，如图 4.8(c) 所示，通过 $H(\mu)$ 和 $\tilde{F}(\mu)$ 相乘我们得到 $F(\mu)$：

> 式(4.3-7)中的 ΔT 抵消了式(4.3-5)中的 $1/\Delta T$。

$$F(\mu)=H(\mu)\tilde{F}(\mu) \tag{4.3-8}$$

一旦得到了 $F(\mu)$，就可以用傅里叶反变换来复原 $f(t)$：

$$f(t)=\int_{-\infty}^{\infty}F(\mu)\mathrm{e}^{\mathrm{j}2\pi\mu t}\mathrm{d}\mu \tag{4.3-9}$$

式(4.3-7)到式(4.3-9)从理论上证明了，以函数包含的最高频率的两倍的速率取样得到的函数的样本，来恢复一个带限函数是可能的。如下一节将要讨论的那样，$f(t)$ 必须是带限的这一要求，通常意味着 $f(t)$ 必须从 $-\infty$ 扩展到 ∞，实际上这一条件并不满足。正如稍后将要看到的那样，限制一个函数的持续时间会妨碍该函数的完美复原，除非情况特殊。

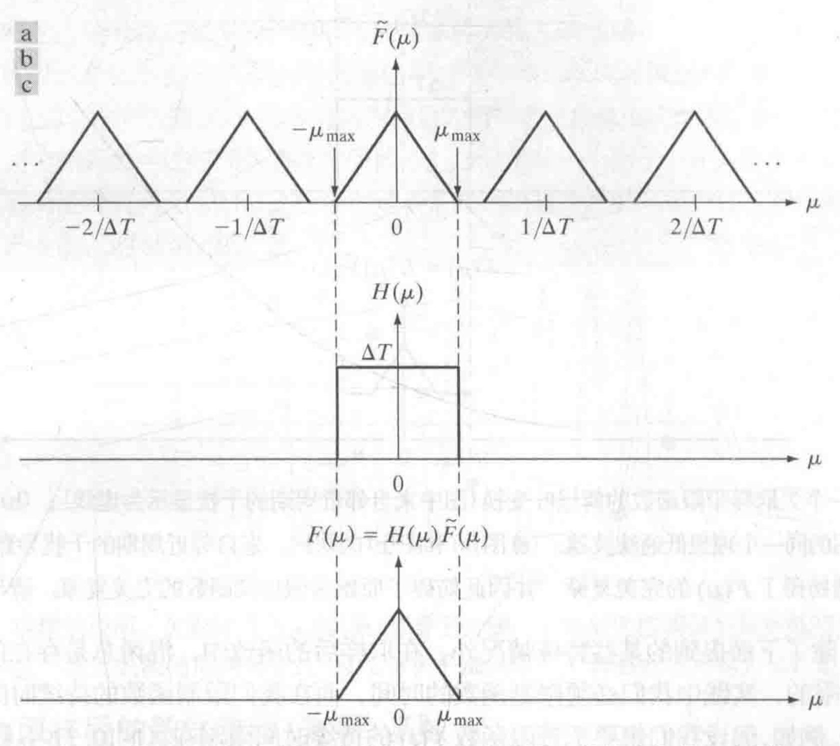


图 4.8　使用一个理想低通滤波器提取一个带限函数的变换的一个周期

函数 $H(\mu)$ 称为低通滤波器，因为它通过频率范围低端的频率，但会消除(滤除)所有较高的频率。它还被称为理想低通滤波器，因为它在幅度上无限快速地过渡(在位置 $-\mu_{\max}$ 处，在 0 和 ΔT 之间，而在位置 $\mu_{\max}$ 处正好相反)，其特性用物理电子元件是不能实现的。我们可以用软件来模拟理想滤波器，但像在 4.7.2 节中解释的那样，尽管这样，也存在限制。在本章后面，我们还会叙述更多有关滤波器的内容。因为滤波器是从函数的取样来恢复(重建)原始函数的手段，因此刚刚讨论的用于此目的的滤波器称为重建滤波器。

4.3.4　混淆

此时一个合理的问题是：如果一个带限函数用低于其最高频率的两倍取样率取样，将会发生什么情况？这相当于前一节中讨论的欠取样情况。图 4.9(a) 与图 4.6(d) 相同，说明了这一情况。以低于奈奎斯特取样率取样的最终效果是周期重叠，且不管使用什么滤波器，都不可能分隔出变

238 ~ 239

换的单个周期。例如，使用图4.9(b)中的理想低通滤波器，将会得到如图4.9(c)所示的一个变换，该变换已被来自邻近周期的频率破坏。然后，反变换会产生 t 的一个破坏了的函数。由函数欠取样导致的这种效果就是我们熟知的频率混淆，或简称为混淆。字面上，混淆是一个过程，在这一过程中，连续函数的高频成分在取样后的函数中被低频“化妆”了。这与通常所用的混淆一词是一致的，意思都是“错认了”。

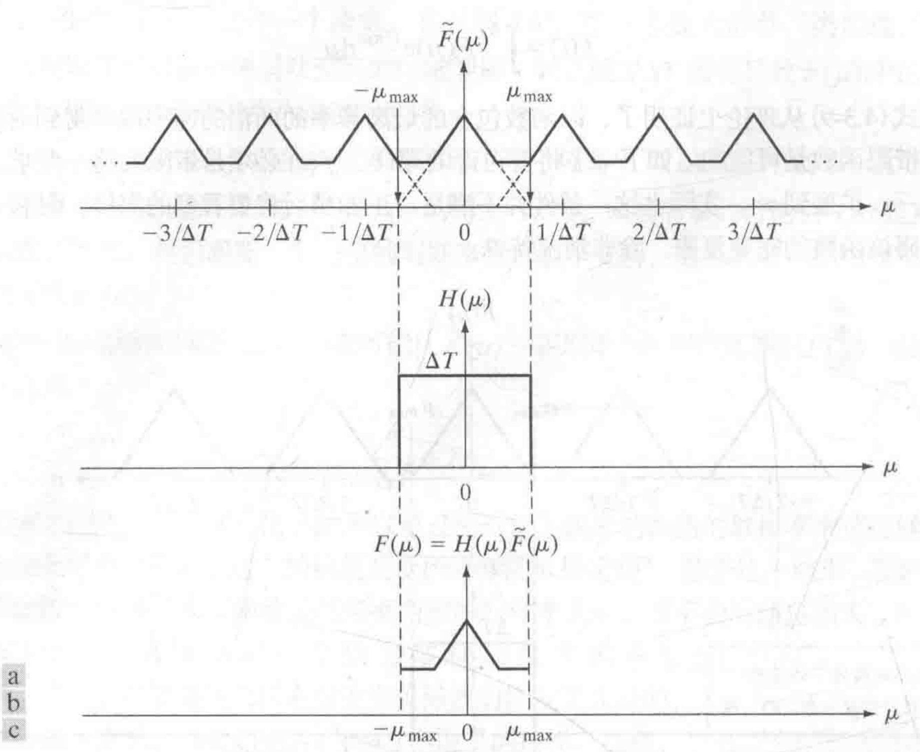


图4.9 (a)一个欠取样带限函数的傅里叶变换(图中来自邻近周期的干扰显示为虚线)；(b)图4.8(b)中所用的同一个理想低通滤波器；(c)图(a)和图(b)的乘积。来自邻近周期的干扰导致了混淆，而混淆妨碍了 $F(\mu)$ 的完美复原，并因此妨碍了原始带限连续函数的完美复原。请与图4.8比较

遗憾的是，除了下面提到的某些特殊情况外，在取样后的函数中，混淆总是存在的，尽管原始取样后的函数是带限的，实践中我们必须限制函数的时间，而在我们限制函数的持续时间时，总会引进无限的频率成分。例如，假设我们想要把带限函数 $f(t)$ 的持续时间限制在区间[0, T]内。我们可以让 $f(t)$ 乘以如下函数来达到：

$$h(t) = \begin{cases} 1, & 0 \le t \le T \\ 0, & \text{其他} \end{cases} \tag{4.3-10}$$

这个函数具有与图4.4(a)相同的基本形状，其变换 $H(\mu)$ 具有无限扩展的频率成分，如图4.4(b)所示。由卷积定理可知，乘积 $h(t)f(t)$ 的变换是这两个函数的变换的卷积。虽然 $f(t)$ 的变换是带限的，将该变换与 $H(\mu)$ 卷积，它包含在另一个函数上滑动一个函数的过程，这就产生了一个频率成分无限扩展的结果。因此，没有有限持续时间的函数是带限的。反过来，带限函数一定从 $-\infty$ 扩展到 ∞ [①]。

根据前一段说明的原因，我们断定，用有限长度的取样和记录工作，混淆将不可避免。在实践中，可以通过平滑输入函数减少高频成分的方法(如对图像采用散焦方法)来降低混淆的影响。

239 ~ 240

① 一个重要的特例是当一个从 $-\infty$ 扩展到 ∞ 的函数是带限的周期函数时。此时，该函数可被截断并且仍是带限的，前提是该截断精确包含整数个周期。单个截断的周期(及函数)可由一组满足取样定理并在截断的区间上获得的离散样本来表示。

这种处理称为抗混淆，它必须在函数被取样之前完成，因为混淆是一个取样问题，而取样问题不能使用计算技术"事后取消"。

例 4.3　混淆。

图4.10显示了混淆的一个经典例子。在两个方向上无限扩展的纯正弦波具有单一频率，因此它很明显是带限的。假定图中的正弦波(现在忽略图中的大点)用式 $\sin(\pi t)$ 来表示，且水平轴对应于时间 t(单位为秒)。函数在 $t=\cdots,-1,0,1,2,3,\cdots$ 处与时间轴相交。

回忆可知，1 个周期/秒定义为 1 Hz。

$\sin(\pi t)$ 的周期 P 是 2 s，其频率是 $1/P$ 或 1/2 周期/秒。根据取样定理，如果取样率 $1/\Delta T$ 超过某一信号最高频率的两倍，我们可以由取样后的一组样本来复原该信号。这意味着取样率大于 1 个样本/秒 $[2\times(1/2)=1]$ 或 $\Delta T<1\,\text{s}$ 时，才能恢复该信号。显然，以准确的两倍频率(1 个样本/秒)的取样率对信号取样，在 $t=\cdots,-1,0,1,2,3,\cdots$ 处得到的样本是 $\cdots\sin(-\pi)$, $\sin(0)$, $\sin(\pi)$, $\sin(2\pi),\cdots$，它们都是零。这说明了取样定理要求取样率超过信号最高频率两倍的原因，就像前面提及的那样。

图4.10中的大黑点是以小于 1 个样本/秒的取样率对信号均匀取样后得到的样本(事实上，样本间的间隔超过了 2 s，它给出的取样率低于 1/2 个样本/秒)。取样后的信号看上去像正弦波，但其频率是原始信号的 1/10。这个频率低于原始连续函数中任何频率的取样后信号，是混淆的一个例子。在只给出图4.10中样本的情形下，混淆的严重程度使得我们无法知道这些样本是不是原始函数的真实描述。如在本章稍后看到的那样，图像中的混淆会产生类似的误导结果。

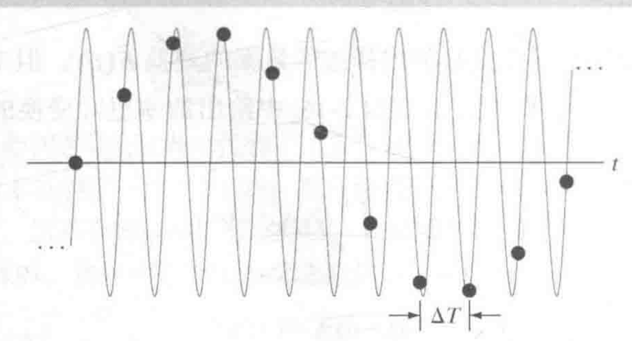


图 4.10　混淆的说明。欠取样后的函数(黑点)看上去像一个频率比连续信号频率低得多的正弦波。该正弦波的周期是 2 s，因此与水平轴的零交叉每秒出现一次，ΔT 是样本间隔

4.3.5　由取样后的数据重建(复原)函数

本节说明由一组样本集合来重建函数实际上可以减少样本间的内插。即使简单地用显示介质显示一幅图像的行为，都要求来自其样本的图像重建。因此，理解取样数据重建的基础是很重要的。卷积是这一理解的核心，因而再度表明了这一概念的重要性。

图 4.8 和式(4.3-8)的讨论概括了一个带限函数使用频率域方法由其样本完美复原的过程。使用卷积定理，我们可以在空间域得到等价的结果。从式(4.3-8)，$F(\mu)=H(\mu)\tilde{F}(\mu)$，可知

$$f(t)=\Im^{-1}\{F(\mu)\}=\Im^{-1}\left\{H(\mu)\tilde{F}(\mu)\right\}=h(t)\bigstar\tilde{f}(t) \tag{4.3-11}$$

式中，最后一步来自式(4.2-21)的卷积定理。可以证明(见习题 4.6)，将式(4.3-1)给出的 $\tilde{f}(t)$ 代入式(4.3-11)，然后使用式(4.2-20)，可导出 $f(t)$ 的如下空间域表达式：

$$f(t)=\sum_{n=-\infty}^{\infty}f(n\Delta T)\operatorname{sinc}\left[(t-n\Delta T)/\Delta T\right] \tag{4.3-12}$$

240
~
242

式中，sinc 函数由式(4.2-19)定义。这个结果并不意外，因为盒状滤波器 $H(\mu)$ 的傅里叶反变换是一个 sinc 函数(见例4.1)。式(4.3-12)表明，完美重建的函数是用取样值加权的 sinc 函数的无限和，并且具有一个重要特性，即重建的函数恒等于在多个 ΔT 的整数增量处的样本值。也就是说，对于任何 $t = k\Delta T$，其中 k 是整数，$f(t)$ 等于第 k 个样本 $f(k\Delta T)$。这来自式(4.3-12)，因为 $\text{sinc}(0) = 1$，并且对于任何其他整数 m，$\text{sinc}(m) = 0$。样本点之间的 $f(t)$ 值是由 sinc 函数的和形成的内插值。

式(4.3-12)要求样本间的内插有无限多项。在实际中，这意味着我们必须找到一种样本间内插有限的近似方法。如我们在 2.4.4 节中讨论的那样，在图像处理中使用的主要内插方法是最近邻法、双线性法和双三次内插法。在 4.5.4 节中，我们将讨论图像内插的效果。

4.4 单变量的离散傅里叶变换(DFT)

本节的主要目的是从基本原理开始推导离散傅里叶变换。迄今为止的内容可视为这些基本原理的基础，因此现在我们适当地推导 DFT 所必需的工具。

4.4.1 由取样后的函数的连续变换得到 DFT

如 4.3.2 节讨论的那样，一个取样过的、带限的、扩展到 $-\infty$ 到 ∞ 范围的函数的傅里叶变换，也是扩展到 $-\infty$ 到 ∞ 范围的连续的周期函数。实践中，我们处理有限数量的样本，因此本节的目的是推导对应于这种样本集合的 DFT。

式(4.3-5)给出了关于原始函数变换的取样过的数据的变换 $\tilde{F}(\mu)$，但未给出采样后函数 $\tilde{f}(t)$ 本身的变换 $\tilde{F}(\mu)$ 的表达式。我们直接由式(4.2-16)中给出的傅里叶变换的定义寻找这样一个表达式：

$$\tilde{F}(\mu) = \int_{-\infty}^{\infty} \tilde{f}(t) e^{-j2\pi\mu t} dt \tag{4.4-1}$$

用式(4.3-1)代替 $\tilde{f}(t)$，得到

$$\begin{aligned} \tilde{F}(\mu) &= \int_{-\infty}^{\infty} \tilde{f}(t) e^{-j2\pi\mu t} dt = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(t)\delta(t - n\Delta T) e^{-j2\pi\mu t} dt = \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} f(t)\delta(t - n\Delta T) e^{-j2\pi\mu t} dt \\ &= \sum_{n=-\infty}^{\infty} f_n e^{-j2\pi\mu n\Delta T} \end{aligned} \tag{4.4-2}$$

式中，最后一步来自式(4.3-2)。虽然 f_n 是离散函数，但如我们由式(4.3-5)所知道的那样，其傅里叶变换 $\tilde{F}(\mu)$ 是周期为 $1/\Delta T$ 的无限周期连续函数。因此，我们需要表征 $\tilde{F}(\mu)$ 的一个周期，而对一个周期取样是 DFT 的基础。

假设我们想要在周期 $\mu = 0$ 到 $\mu = 1/\Delta T$ 之间得到 $\tilde{F}(\mu)$ 的 M 个等间距样本。这可通过在如下频率处取样得到：

$$\mu = \frac{m}{M\Delta T}, \quad m = 0,1,2,\cdots,M-1 \tag{4.4-3}$$

把 μ 的这一结果代入式(4.4-2)，并令 F_m 表示得到的结果，则有

242 ~ 243

$$F_m = \sum_{n=0}^{M-1} f_n e^{-j2\pi mn/M}, \quad m = 0,1,2,\cdots,M-1 \tag{4.4-4}$$

这个表达式就是我们寻找的离散傅里叶变换[1]。给定一个由 $f(t)$ 的 M 个样本组成的集合 $\{f_n\}$，式(4.4-4)得出一个与输入样本集合离散傅里叶变换相对应的 M 个复离散值的样本集合 $\{F_m\}$。反之，给定 $\{F_m\}$，我们可以用傅里叶反变换(IDFT)复原样本集 $\{f_n\}$：

$$f_n = \frac{1}{M}\sum_{m=0}^{M-1} F_m e^{j2\pi mn/M}, \quad n = 0,1,2,\cdots,M-1 \tag{4.4-5}$$

不难证明(见习题 4.8)，把式(4.4-5)的 f_n 代入式(4.4-4)可得出 $F_m \equiv F_m$。类似地，把式(4.4-4)中的 F_m 代入式(4.4-5)得到 $f_n \equiv f_n$。这意味着式(4.4-4)和式(4.4-5)组成了一个离散傅里叶变换对。此外，这些恒等式指出，对于任何其值有限的样本集合，正、反傅里叶变换都是存在的。注意，这两个表达式不明确地依赖于取样间隔 ΔT，也不依赖于式(4.4-3)中的频率间隔。因此，离散傅里叶变换对适用于任何均匀取样的有限离散样本集。

在前面的阐述中，我们使用 m 和 n 来表示离散变量，因为人们在推导中历来都是这样做的。然而，特别是在二维情况下，使用 x 和 y 表示图像坐标变量并使用 u 和 v 表示频率变量更为直观，在这里，这些变量可理解为整数。这样，式(4.4-4)和式(4.4-5)就变为

$$F(u) = \sum_{x=0}^{M-1} f(x)e^{-j2\pi ux/M}, \quad u = 0,1,2,\cdots,M-1 \tag{4.4-6}$$

和

$$f(x) = \frac{1}{M}\sum_{u=0}^{M-1} F(u)e^{j2\pi ux/M}, \quad x = 0,1,2,\cdots,M-1 \tag{4.4-7}$$

为简便起见，我们在上式中使用函数表示代替了下标[2]。显然，$F(u) \equiv F_m$ 且 $f(x) \equiv f_n$。从这一点出发，我们用式(4.4-6)和式(4.4-7)表示一维 DFT 对。某些作者通过在式(4.4-6)中包含 1/M 项来代替我们在式(4.4-7)中所示的方法。这不会影响两个公式形成一个傅里叶变换对的证明。

可以证明(见习题 4.9)，傅里叶正变换和反变换都是无限周期的，其周期为 M，即

$$F(u) = F(u + kM) \tag{4.4-8}$$

和

$$f(x) = f(x + kM) \tag{4.4-9}$$

式中，k 是整数。

> 考虑到对连续函数取样的结果不是周期的，离散函数 $f(x)$是周期函数的原因并不明显。解释这一点的一种非正式方法是，记住一个周期 DFT 中的取样结果。反 DFT 的结果 $f(x)$对于 DFT 对的存在来说必须是周期的。

式(4.2-20)中的卷积的离散等价表示是

$$f(x)\bigstar h(x) = \sum_{m=0}^{M-1} f(m)h(x-m) \tag{4.4-10}$$

$x = 0, 1, 2, \cdots, M-1$。因为在前面的公式中，函数是周期的，所以它们的卷积也是周期的。式(4.4-10)给出了周期卷积的一个周期。由于这一原因，在这个公式中，其内在处理通常称为*循环卷积*，并且直接导致 DFT 和它的反变换的周期性。这与我们在 3.4.2 节中研究的卷积形成了对比，在该卷积中，位移

① 注意图 4.6(b)中的区间[0, 1/ΔT]覆盖了该变换的两个背靠背的半周期。这意味着要求重新对 F_m 中的数据排序，以便得到一个周期内频率从最低到最高的样本。这是为符号表示的方便性而必须付出的代价，即在 $m = 0,1,\cdots,M-1$ 处采样，而不使用原点两侧的样本，因为后者要求使用负号。4.6.3 节将讨论对变换数据排序的过程。

② 使用 t 表示连续空间变量及使用 μ 表示连续频率变量时，我们一直非常小心。从现在起，我们将使用 x 和 u 来分别表示一维离散空间变量和一维频率变量。在处理二维函数时，我们将使用(t, z)和(μ, ν)来分别表示连续空间变量和频率域变量。类似地，我们将使用(x, y)和(u, v)来表示与它们对应的离散变量。

244 ~ 245

值 x 由一个函数完全滑过另一个函数的要求来决定，且像在循环卷积中那样区间$[0, M-1]$是不固定的。我们将在 4.6.3 节和图4.28 中讨论这一区别及其意义。

最后，我们指出，式(4.2-21)和式(4.2-22)中给出的卷积定理还可用于离散变量(见习题 4.10)。

4.4.2 取样和频率间隔间的关系

如果 $f(x)$ 由函数 $f(t)$ 以 ΔT 为单位间隔取样后的 M 个样本组成，则包含集合 $\{f(x)\}$，$x = 0, 1, 2, \cdots, M-1$的记录的持续时间是

$$T = M\Delta T \tag{4.4-11}$$

离散频率域中的相应间隔 Δu 来自式(4.4-3)：

$$\Delta u = \frac{1}{M\Delta T} = \frac{1}{T} \tag{4.4-12}$$

由 DFT 的 M 个成分跨越的整个频率范围是

$$\Omega = M\Delta u = \frac{1}{\Delta T} \tag{4.4-13}$$

这样，由式(4.4-12)和式(4.4-13)就可以看出，DFT 的频率分辨率 Δu 取决于连续函数 $f(t)$被取样的持续时间 T，并且 DFT 跨越的频率范围取决于取样间隔 ΔT 。很明显，两个表达式给出了 T 和 ΔT 的反转关系。

例 4.4 计算 DFT 的机理。

图4.11(a)显示了连续函数 $f(t)$ 以 ΔT 为单位间隔取样后的 4 个样本。图4.11(b)显示了 x 域中的取样值。注意，x 的值是 0, 1, 2 和 3，它指出我们可以提供 $f(t)$ 的任意 4 个样本。

由式(4.4-6)有

$$F(0) = \sum_{x=0}^{3} f(x) = [f(0) + f(1) + f(2) + f(3)] = 1 + 2 + 4 + 4 = 11$$

$F(u)$的下一个值是

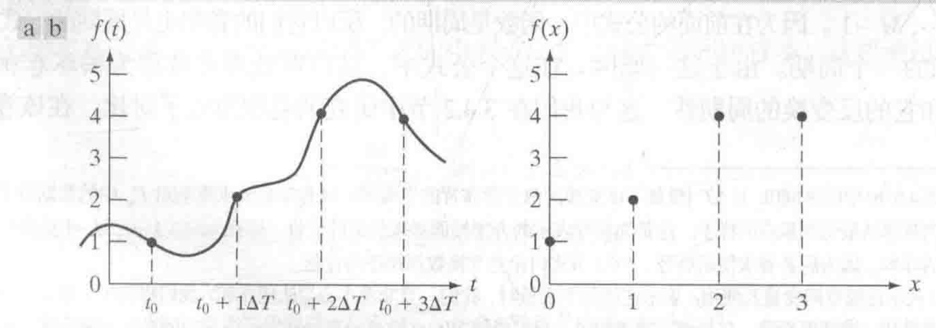
$$F(1) = \sum_{x=0}^{3} f(x)\mathrm{e}^{-\mathrm{j}2\pi(1)x/4} = 1\mathrm{e}^{0} + 2\mathrm{e}^{-\mathrm{j}\pi/2} + 4\mathrm{e}^{-\mathrm{j}\pi} + 4\mathrm{e}^{-\mathrm{j}3\pi/2} = -3 + 2\mathrm{j}$$

类似地，有 $F(2) = -(1 + 0\mathrm{j})$ 和 $F(3) = -(3 + 2\mathrm{j})$ 。注意观察，在计算 $F(u)$ 的每一项时用到了 $f(x)$ 的所有值。

如果给定的是 $F(u)$，并且要求计算它的反变换，那么我们可以用相同的方法进行处理，但使用的是反变换。例如，

$$f(0) = \frac{1}{4}\sum_{u=0}^{3} F(u)\mathrm{e}^{\mathrm{j}2\pi u(0)} = \frac{1}{4}\sum_{u=0}^{3} F(u) = \frac{1}{4}[11 - 3 + 2\mathrm{j} - 1 - 3 - 2\mathrm{j}] = \frac{1}{4}[4] = 1$$

这与图4.11(b)一致。$f(x)$的其他值可以用相同的方法得到。



245 ~ 246

图 4.11 (a)一个函数；(b)该函数在 x 域中的样本。在图(a)中，t 是一个连续变量；在图(b)中，x 代表整数值

4.5　两个变量的函数的扩展

本节我们把 4.2 节到 4.4 节介绍的概念扩展到两个变量的情形。

4.5.1　二维冲激及其取样特性

两个连续变量 t 和 z 的冲激 $\delta(t,z)$ 定义为如式(4.2-8)所示的形式：

$$\delta(t,z)=\begin{cases}\infty, & t=z=0\\ 0, & \text{其他}\end{cases} \tag{4.5-1a}$$

和

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\delta(t,z)\mathrm{d}t\,\mathrm{d}z=1 \tag{4.5-1b}$$

如在一维情况那样，二维冲激在积分下呈现了取样特性：

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(t,z)\delta(t,z)\mathrm{d}t\,\mathrm{d}z=f(0,0) \tag{4.5-2}$$

或者，更一般地，对位于坐标 (t_0,z_0) 处的冲激，有

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(t,z)\delta(t-t_0,z-z_0)\mathrm{d}t\,\mathrm{d}z=f(t_0,z_0) \tag{4.5-3}$$

如以前那样，我们看到，取样特性在冲激所在的位置产生函数 $f(t,z)$的值。

对于离散变量 x 和 y，二维离散冲激定义为

$$\delta(x,y)=\begin{cases}1, & x=y=0\\ 0, & \text{其他}\end{cases} \tag{4.5-4}$$

其取样特性为

$$\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}f(x,y)\delta(x,y)=f(0,0) \tag{4.5-5}$$

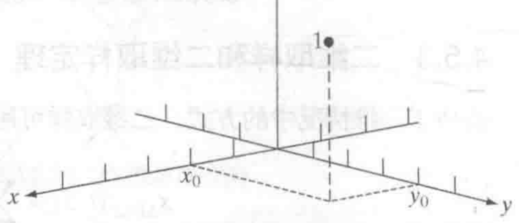


图 4.12　二维单位离散冲激。变量 x 和 y 是离散的，且 δ 的值在除坐标 (x_0,y_0) 外的任何位置都为 0

式中，$f(x,y)$是离散变量 x 和 y 的函数。对于一个位于坐标 (x_0,y_0) 外的冲激(见图 4.12)，其取样特性为

$$\sum_{x=-\infty}^{\infty}\sum_{y=-\infty}^{\infty}f(x,y)\delta(x-x_0,y-y_0)=f(x_0,y_0) \tag{4.5-6}$$

如以前那样，离散冲激的取样特性在该冲激所在的位置产生离散函数 $f(x,y)$的值。

4.5.2　二维连续傅里叶变换对

令 $f(t,z)$ 是两个连续变量 t 和 z 的连续函数。则其二维连续傅里叶变换对由以下两个表达式给出：

$$F(\mu,\nu)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(t,z)\mathrm{e}^{-\mathrm{j}2\pi(\mu t+\nu z)}\mathrm{d}t\,\mathrm{d}z \tag{4.5-7}$$

和

$$f(t,z)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(\mu,\nu)\mathrm{e}^{\mathrm{j}2\pi(\mu t+\nu z)}\mathrm{d}\mu\,\mathrm{d}\nu \tag{4.5-8}$$

247 ~ 248

式中，μ 和 v 是频率变量。当涉及图像时，t 和 z 解释为连续空间变量。就像一维情况那样，变量 μ 和 v 的域定义了连续频率域。

例 4.5　得到一个简单函数的二维傅里叶变换。

图 4.13 显示了一个模仿例 4.1 中一维情况的二维函数。按照类似于例 4.1 中使用的过程给出结果

$$F(\mu,v)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(t,z)\mathrm{e}^{-\mathrm{j}2\pi(\mu t+vz)}\mathrm{d}t\,\mathrm{d}z=\int_{-T/2}^{T/2}\int_{-Z/2}^{Z/2}A\mathrm{e}^{-\mathrm{j}2\pi(\mu t+vz)}\mathrm{d}t\,\mathrm{d}z=ATZ\left[\frac{\sin(\pi\mu T)}{(\pi\mu T)}\right]\left[\frac{\sin(\pi vZ)}{(\pi vZ)}\right]$$

其幅度(谱)由下面的表达式给出：

$$|F(\mu,v)|=ATZ\left|\frac{\sin(\pi\mu T)}{(\pi\mu T)}\right|\left|\frac{\sin(\pi vZ)}{(\pi vZ)}\right|$$

图4.13(b)显示了关于原点的谱的一部分。正如一维的情况那样，在谱中零的位置与 T 和 Z 的值成反比。这样，T 和 Z 越大，谱变得越“紧凑”，反之亦然。

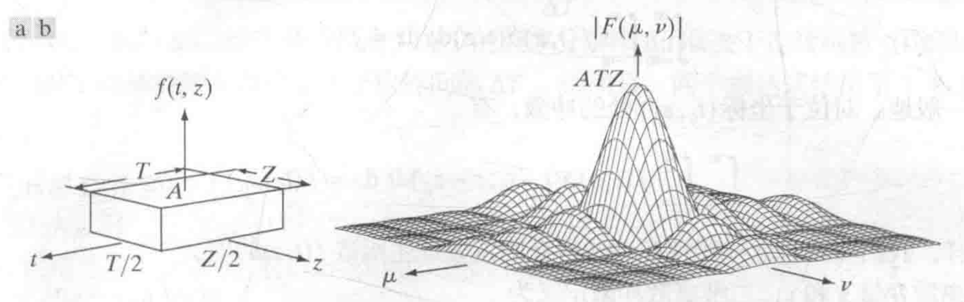


图 4.13　(a)一个二维函数；(b)该二维函数的谱的一部分(未按比例画出)，该块沿 t 轴越长，则沿 μ 轴的谱更“紧凑”。请与图 4.4 比较

4.5.3　二维取样和二维取样定理

类似于一维情况中的方式，二维取样可用取样函数(二维冲激串)建模：

$$s_{\Delta T\Delta Z}(t,z)=\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\delta(t-m\Delta T,z-n\Delta Z)\tag{4.5-9}$$

式中，ΔT 和 ΔZ 是连续函数 $f(t,z)$ 沿 t 轴和 z 轴的样本间的间隔。式(4.5-9)描述了沿着两个轴无限扩展的周期冲激的集合(见图4.14)。如图4.5中说明的一维情况那样，用 $s_{\Delta T\Delta Z}(t,z)$ 乘 $f(t,z)$ 可得到取样后的函数。

如果由区间 $[-\mu_{\max},\mu_{\max}]$ 和区间 $[-v_{\max},v_{\max}]$ 建立的矩形之外的傅里叶变换是零，则函数 $f(t,z)$ 称为带限函数；即

$$F(\mu,v)=0,\quad |\mu|\geq \mu_{\max}\text{且}|v|\geq v_{\max}\tag{4.5-10}$$

图 4.14　二维冲激串

二维取样定理表明，如果取样间隔满足

$$\Delta T<\frac{1}{2\mu_{\max}}\tag{4.5-11}$$

248~249

和

$$\Delta Z < \frac{1}{2v_{\max}} \tag{4.5-12}$$

或者关于取样率的表达，

$$\frac{1}{\Delta T} > 2\mu_{\max} \tag{4.5-13}$$

和

$$\frac{1}{\Delta Z} > 2v_{\max} \tag{4.5-14}$$

则连续带限函数 $f(t,z)$ 可以由其一组样本无误地恢复。另一种叙述方法是：我们说，如果一个二维带限连续函数在 μ 和 v 两个方向上由以大于该函数最高频率两倍的取样率取样获得的样本表示，则没有信息丢失。

图4.15 显示了图4.6(b)和(d)的二维等效图示。一个理想的二维盒状滤波器有图 4.13(a)说明的形式。图4.15(a)中的虚线部分说明了滤波器的位置，以便分离从其样本重建带限函数的必要变换的单个周期，正如4.3.3 节那样。由 4.3.4 节我们知道，如果函数欠取样，则周期将会重叠，并且如图4.15(b)所示，分离出单个周期是不可能的。在这样的条件下，将产生混淆。

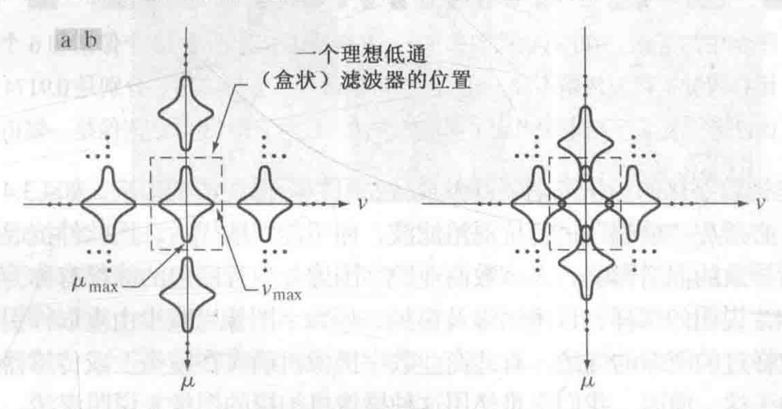


图 4.15　(a)过取样和(b)欠取样带限函数的二维傅里叶变换

4.5.4　图像中的混淆

本节把混淆的概念扩展到图像，并讨论一些与图像取样和重取样有关的问题。

从一维混淆展开

如一维情况下那样，如果只在两个坐标方向上无限扩展，二维连续变量 t 和 z 的连续函数通常可以被带限。限制一个函数时间的非常动作，将引起在频率域无限扩展的有害频率成分，如 4.3.4 节所解释的那样。因为我们不可能对一个函数无限地取样，所以，正像在取样后的一维函数中所存在的那样，在数字图像中混淆总是存在的。在图像中，存在两种主要的混淆现象：空间混淆和时间混淆。如4.3.4 节讨论的那样，空间混淆是由欠取样造成的。时间混淆与图像序列中图像间的时间间隔有关。时间混淆最常见的例子之一是“车轮”效应，即在序列图像(如电影)中具有轮辐的车轮出现倒转的现象。这是由序列中的帧率低于车轮转动的速度引起的。

本章重点关注空间混淆。图像中空间混淆的主要表现形式是人为引入的缺陷，如线状特征中的锯齿、伪高光及原图像中不存在的模式。下面的例子说明了图像中的混淆。

249 ~ 251

例 4.6 图像中的混淆。

假定我们有一个很完美的成像系统，没有噪声，并能产生看上去很精细的数字图像，但它能采用的取样数是固定的 96 × 96 个像素。如果我们用该系统去数字化棋盘图案，那么它能解决多达 96 × 96 个大小为 1×1 像素的方格图案。在这种限制下，结果图像中的每个像素将对应图案中的一个方格。我们感兴趣的是考察当细节(棋盘方格的大小)小于一个摄像机像素时会发生什么事情；也就是说，当成像系统要数字化大于 96 × 96 个方格的棋盘图案时，会发生什么情况。

> 这个例子应切合实际地重建。当我们将计算机生成的模型和向量图转换为数字图像时，在无噪声、无失真的条件下对“完美”的场景取样是很常见的。

图4.16(a)和(b)显示了对棋盘取样的结果，方格的边长分别为 16 像素和 6 像素。这些结果正如所预料的那样。然而，当方格的大小减小到稍小于一个摄像机像素时，就会产生如图 4.16(c)所示的严重混淆的图像。最后，将方格的边长减小到稍小于 0.5 个像素时，会产生如图4.16(b)所示的图像。在这种情况下，混淆的结果看上去还像是一个正常的棋盘图案。事实上，该图像是通过对其方格边长为 12 像素的棋盘图像取样得到的。最后一幅图像是较好的提醒，即混淆会产生令人完全误解的结果。

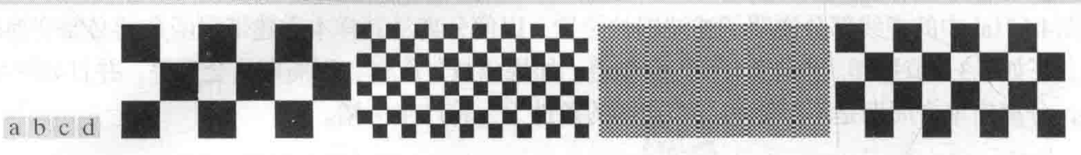


图 4.16 图像中的混淆：在图(a)和图(b)中，方格的边长分别是 16 个像素和 6 个像素，混淆在视觉上可以忽略不计。在图(c)和图(d)中，方格的边长分别是 0.9174 个像素和 0.4798 个像素，结果中出现了明显的混淆。注意，图(d)看起来像是一幅正常的图像

通过稍微散焦被数字化的场景来削弱高频成分，可以降低混淆的影响。如4.3.4 节解释的那样，在图像被取样之前，必须在“前端”进行抗混淆滤波，而不能在取样后，用软件抗混淆滤波器来降低由于违反取样定理所导致的混淆影响。大多数商业数字图像处理程序包的确都有称为反混淆的功能。然而，如例4.7 和例4.8 说明的那样，该术语涉及模糊一幅数字图像以减少由重取样引起的附加混淆，不能用于减少原始取样过的图像的混淆。有些商业数字摄像机确实在镜头上或传感器本身的表面上嵌入了反混淆滤波。由于这一原因，我们很难使用这种摄像机拍摄的图像来说明混淆。

图像内插和重取样

就像在一维情形中那样，由样本集合完美地重建一个带限图像函数时，要求在空间域用 sinc 函数做二维卷积。如4.3.5 节解释的那样，这种理论上的完美重建要求使用无限求和来内插，而在实践中则迫使我们不得不寻求某种近似。在图像处理中，二维内插最普通的应用之一是调整图像的大小(放大和缩小)。放大可视为过取样，缩小则可视为欠取样。这两种操作和前面几节讨论的取样概念之间的主要区别是，放大和缩小用于数字图像。

2.4.4 节中已解释过内插。这里我们的兴趣在于说明最近邻、双线性和双三次内插的性能。本节给出一些关于取样和反混淆问题的例子。当我们想要以整数倍增大一幅图像的尺寸时，与过取样相关的最近邻内插的一种特殊情况是，采用像素复制的放大操作。例如，将一幅图像放大两倍，我们可以先复制每一列。这会在水平方向上将图像尺寸放大一倍。然后，我们复制这幅放大后的图像的每一行，在垂直方向上将图像尺寸放大一倍。相同的步骤可用于将图像放大任意整数倍。每个像素的灰度级分配，由新位置是老位置的精确复制这一事实来预先确定。

251 ~ 253

采用类似于放大图像的方法，同样可以缩小图像。欠取样是通过行-列删除来实现的(例如，要将图像缩小一半，可每隔一行和一列来删除一

> 不使用带限模糊方法对图像重取样的过程称为抽取。

行和一列)。我们可以用 2.4.4 节的放大网格模拟来形象化非整数参数收缩的概念，只是我们现在扩张该网格拟合原图像，进行灰度内插，然后再把网格缩小为规定的尺寸。为减少混淆，最好在缩小图像之前稍微模糊一下图像(4.8 节将讨论频率域模糊)。一种替代技术是对原始场景超取样，然后采用行-列删除法来减小(重采样)其尺寸。这可以得到比平滑更清晰的结果，但是，很显然，它要求存取原始场景。如果我们不能访问原始场景(实际中这很常见)，那么超取样就不是一个好的选择。

例 4.7　重取样后的图像中的混淆的说明。

一幅数字图像缩小后，混淆效应通常会变得更严重。图4.17(a)是为说明混淆的影响而故意创建的一幅图像(注意，外衣上所有隔开的较细平行线都已被物体磨损)。图4.17(a)中没有不好的人为缺陷，这表明所用的初始取样率对防止可见的混淆是有效的。在图4.17(b)中，图像用行-列删除法缩小为原尺寸的 50%。在该图像中，混淆的影响十分明显(作为例子，可参见膝盖周围的区域)。连续图像的反混淆滤波的数字等效操作是，在对图像进行重取样之前平滑图像，进而衰减数字图像的高频成分。图4.17(c)显示了在缩小其尺寸之前，使用 3 × 3 均值滤波器(见3.5 节)平滑图像的结果。它对图 4.17(b)的改进效果很明显。为了方便比较，采用像素复制方法将图 4.17(b)和图 4.17(c)放大到原始尺寸。





a b c

图 4.17　重取样后的图像上的混淆的说明：(a)带有视觉上可以忽略的混淆的数字图像；(b)采用行-列删除法将图像尺寸缩小到其原始尺寸 50%后的结果。混淆清晰可见；(c)在调整图像大小之前，使用一个 3×3 均值滤波器来模糊(a)中图像的结果。该图像比图(b)更模糊，但混淆不再那样显眼(原图像由加州大学圣巴巴拉分校信号压缩实验室提供)

当我们处理具有很强边缘内容的图像时，混淆的影响看起来就像是块状图像成分，这称为锯齿。下面的例子说明了这一现象。

例 4.8　图像缩小中锯齿现象的说明。

图4.18(a)显示了一幅计算机生成场景的 1024 × 1024 数字图像，图像中混淆可以忽略。为了使混淆效果(这种情况下是锯齿)更清楚，图 4.18(b)显示了经过如下操作后的图像：将图(a)缩小到 75%，即 256 × 256 大小，然后使用双线性内插法和像素复制法，将该图像复原到原始尺寸。就像例 4.7 中那样，在对图像重取样之前，平滑该图像可使混淆的影响不太令人讨厌。图 4.18(c)是在减小图像尺寸之前使用 5 × 5 均值滤波器的结果。正如这幅图像显示的那样，锯齿明显减少了。采用生成图 4.18(c)的相同方法，图 4.18(c)同样经过了图像缩小与复原为其原始尺寸的过程。

例 4.9　图像放大中锯齿现象的说明。

在前两个例子中，我们使用像素复制法来放大较小的取样后图像。如图4.19说明的那样，通常这并不是一种首选的方法。图4.19(a)显示了一幅放大为 1024×1024 像素的图像，它是从图 4.18(a)所示图像中心裁出 256 × 256 像素部分，通过像素复制的方法产生的。注意“斑驳”的边缘。图4.19(b)所示放大后的图像是由相同的 256×256 像素部分经双线性内插法产生的。在该结果中，边缘相当平滑。例如，瓶颈的边缘和大棋盘方格不像图 4.19(b)中那么斑驳，而与图 4.19(a)相近。

253 ~ 255

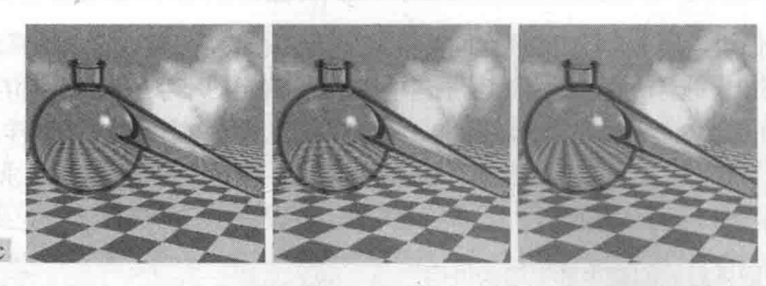


图 4.18　锯齿的说明：(a)一幅计算机生成场景的 1024×1024 数字图像，其混淆可以忽略；(b)减小(a)的原始尺寸到 25%，再用双线性内插恢复原始尺寸的结果。(c)减小图像尺寸到 25%之前，用 5×5 的均值滤波模糊图像，然后再用双线性内插的结果(原图像由 D. P. Mitchell, Mental Landscape, LLC 提供)

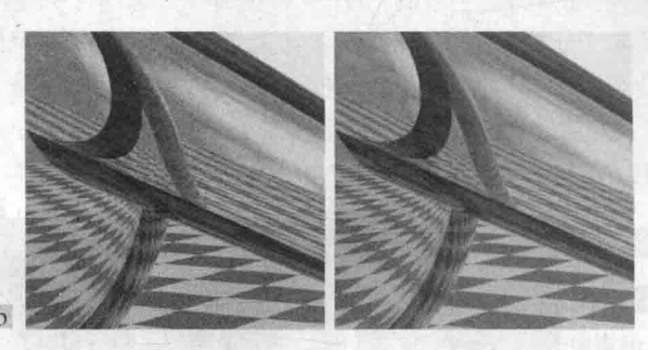


图 4.19　图像放大：(a)一幅大小为 1024×1024 像素的数字图像，该图像是通过对图 4.18(a)中间裁取的 256×256 部分使用像素复制法产生的；(b)使用双线性内插法生成的图像，图像中的锯齿明显减少

莫尔(波纹)模式

在结束本节之前，我们考察一下另外一种人为缺陷，称之为莫尔(波纹)模式[①]，它们有时是使用周期或近似周期成分对场景取样产生的。在光学领域，莫尔模式指的是在两个近似等间隔的光栅之间产生的差拍模式。这些差拍模式通常每天都会发生。例如，防虫纱窗重叠在一起时所产生的干涉现象。在数字图像处理中扫描纸质印刷物如报纸和杂志时，或在具有周期成分的图像中，如果它的间隔与取样的间隔可比时，通常都会出现这一问题。与取样造成的人为缺陷相比，莫尔模式更为常见。例如，图 4.20 显示了莫尔(波纹)效应，它们是油墨画，并未被数字化。分离情况下，这些模式都很干净，且没有干扰。但在将一种模式叠加到另一种模式上时，会产生一幅其频率在原始模式中均不出现的差拍模式。注意，在特殊情况下，由两个点的模式所产生的莫尔(波纹)效应，是下面的讨论很感兴趣的效应。

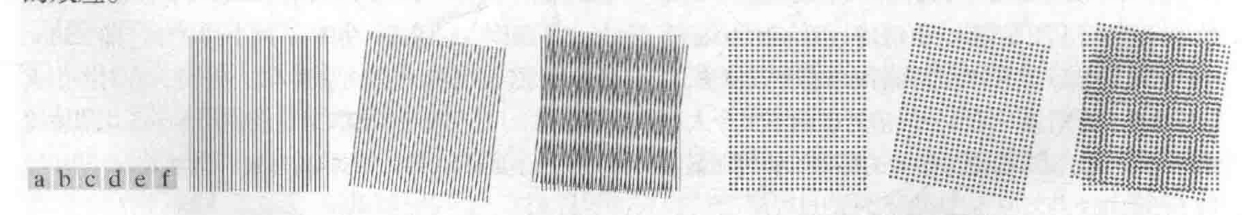


图 4.20　莫尔(波纹)效应的例子。它们是油墨画而不是数字化后的模式。将一个模式叠加在另一个模式上在数学上等价于两个模式相乘

报纸和其他印刷材料使用所谓的半色调点，它们是黑色圆点或椭圆点，其大小和变化的连接方式用于模拟灰色调。通常，下面的数

> 彩色打印使用红点、绿点和蓝点来产生人眼可以识别的连续颜色。

254 ~ 256

① 术语莫尔(moiré)是一个法语单词而非人名，它源自首先注意到某些织物上出现干涉图样的织布者；该术语是由单词 mohair(马海毛，一种使用安哥拉山羊毛织成的布)演化而来的。

值很典型：报纸用每英寸 75 点(简称 dpi)的分辨率印刷，杂志用 133 dpi 印刷，高质量小册子用 175 dpi 印刷。图4.21 显示了一幅报纸图像以 75 dpi 取样时产生的现象。取样格(垂直方向和水平方向)和报纸图像上点的模式(± 45°方向)相互作用，产生使得图像看起来像是布满斑点的均匀莫尔模式(4.10.2 节将讨论降低莫尔干扰模式的一种技术)。

作为相关的兴趣点，图 4.22 显示了一幅以 400 dpi 取样以防止莫尔(波纹)效应的报纸图像。目标左眼周围区域的放大图像说明了如何利用中间色点产生灰度的深浅。点的大小与图像灰度成反比。在亮区域，点较小，或完全没有点(例如，见眼白部分)。在浅灰区域，点较大，如眼睛下部所示。在较暗的区域，当点的大小超过某个指定值(通常为 50%)时，允许点沿两个指定方向连接起来形成一个互相连接的网格(见眼睛的左侧)。在某些情况下，点仅沿某个方向连接，如眉毛下面的右上方区域。

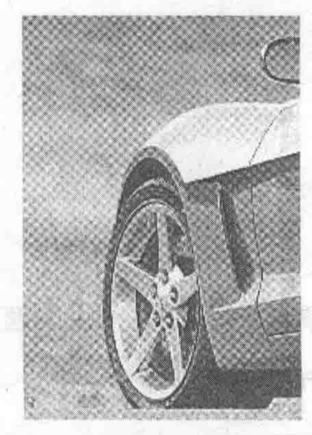


图4.21　以 75 dpi 取样、大小为 246×168 像素的报纸图像，显示了莫尔模式。这幅图像中的莫尔模式是在±45°方向的半色调点与南北方向的取样网格之间产生的干扰

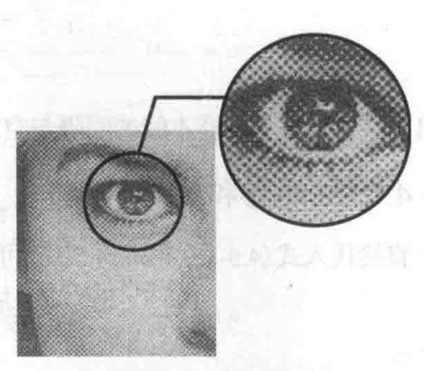


图 4.22　一幅报纸图像及其局部放大图像，局部放大图像说明了如何安排半色调点以渲染灰度的色调

4.5.5　二维离散傅里叶变换及其反变换

类似于 4.3 节和 4.4 节中的推导过程，我们可得到下面的二维傅里叶变换(DFT)：

$$F(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)} \tag{4.5-15}$$

式中，$f(x,y)$ 是大小为 $M\times N$ 的数字图像。与一维中的情况一样，式(4.5-15)必须对离散变量 u 和 v 在 $u=0,1,2,\cdots,M-1$ 和 $v=0,1,2,\cdots,N-1$ 范围内求值[①]。

> 有文献中，我们有时会在 DFT 而不是 IDFT 的前面看到常数 1/MN。这时，该常数表示为 $1/\sqrt{MN}$ 并包含在正变换和反变换的前面，从而创建一个更为对称的变换对。这些公式都是正确的，但前提条件是要保持一致。

给出变换 $F(u,v)$，我们可以使用傅里叶反变换(IDFT)得到 $f(x,y)$：

$$f(x,y)=\frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)\mathrm{e}^{\mathrm{j}2\pi(ux/M+vy/N)} \tag{4.5-16}$$

式中，$x=0,1,2,\cdots,M-1$，$y=0,1,2,\cdots,N-1$。式(4.5-15)和式(4.5-16)构成了二维离散傅里叶变换对。本章的其余部分以这两个公式的特性及它们在频率域图像滤波的应用为基础。

① 如 4.4.1 节中提及的那样，记住，在本章中我们分别使用 (t,z) 和 (μ,ν) 来表示二维连续空间变量和频率域变量。在二维离散情况下，我们使用 (x,y) 表示空间变量，而使用 (u,v) 表示频率域变量。

4.6 二维离散傅里叶变换的一些性质

本节介绍二维离散傅里叶变换及其反变换的一些性质。

4.6.1 空间和频率间隔的关系

空间取样和相应频率域间隔的关系如4.4.2节中解释的那样。假定对连续函数 $f(t,z)$ 取样生成了一幅数字图像 $f(x,y)$，它由分别在 t 和 z 方向所取的 $M\times N$ 个样点组成。令 ΔT 和 ΔZ 表示样本间的间隔(见图4.14)。那么，相应离散频率域变量间的间隔分别由

$$\Delta u = \frac{1}{M\Delta T} \tag{4.6-1}$$

和

$$\Delta v = \frac{1}{N\Delta Z} \tag{4.6-2}$$

给出。注意，频率域样本间的间隔与空间样本间的间距及样本数成反比。

4.6.2 平移和旋转

直接代入式(4.5-15)和式(4.5-16)可以证明傅里叶变换对满足下列平移特性(见习题4.16)：

$$f(x,y)\mathrm{e}^{\mathrm{j}2\pi(u_0x/M+v_0y/N)} \Leftrightarrow F(u-u_0,v-v_0) \tag{4.6-3}$$

和

$$f(x-x_0,y-y_0) \Leftrightarrow F(u,v)\mathrm{e}^{-\mathrm{j}2\pi(x_0u/M+y_0v/N)} \tag{4.6-4}$$

也就是说，用指数项乘以 $f(x,y)$ 将使DFT的原点移到点 (u_0,v_0)；反之，用负指数项乘以 $F(u,v)$ 将使 $f(x,y)$ 的原点移到点 (x_0,y_0)。如例4.13中说明的那样，平移不影响 $F(u,v)$ 的幅度(谱)。

使用极坐标

$$x = r\cos\theta,\quad y = r\sin\theta,\quad u = \omega\cos\varphi,\quad v = \omega\sin\varphi$$

可得到下列变换对：

$$f(r,\theta+\theta_0) \Leftrightarrow F(\omega,\varphi+\varphi_0) \tag{4.6-5}$$

它指出，若 $f(x,y)$ 旋转 θ_0 角度，则 $F(u,v)$ 也旋转相同的角度。反之，若 $F(u,v)$ 旋转一个角度，$f(x,y)$ 也旋转相同的角度。

4.6.3 周期性

如在一维情况中那样，二维傅里叶变换及其反变换在 u 方向和 v 方向是无限周期的，即

$$F(u,v) = F(u+k_1M,v) = F(u,v+k_2N) = F(u+k_1M,v+k_2N) \tag{4.6-6}$$

和

$$f(x,y) = f(x+k_1M,y) = f(x,y+k_2N) = f(x+k_1M,y+k_2N) \tag{4.6-7}$$

式中，k_1 和 k_2 是整数。

变换及其反变换的周期性在基于DFT算法的实现上是重要的特性。考虑图4.23(a)中的一维谱。如4.4.1节描述的那样，在区间[0, M − 1]上，变换数据由两个在点M/2处相会且背靠背的半个周期组成。针对显示和滤波目的，在该区间中有变换的一个完整周期会更为方便，因为完整周期中数据是连续的，如图4.23(b)那样。由式(4.6-3)可得

258 ~ 259

$$f(x)e^{j2\pi(u_0x/M)} \Leftrightarrow F(u-u_0)$$

换句话说，用一个指数项乘以 $f(x)$ 将位于原点的数据 $F(0)$移动到位置u_0。若令 $u_0 = M/2$，则指数项变为 $e^{j\pi x}$，因为 x 是整数，故它等于 $(-1)^x$。在这种情况下，

$$f(x)(-1)^x \Leftrightarrow F(u-M/2)$$

即用 $(-1)^x$ 乘以 $f(x)$ 将位于原点的数据 $F(0)$移动到区间$[0, M-1]$的中心位置，像期望的那样，它对应于图 4.23(b)。

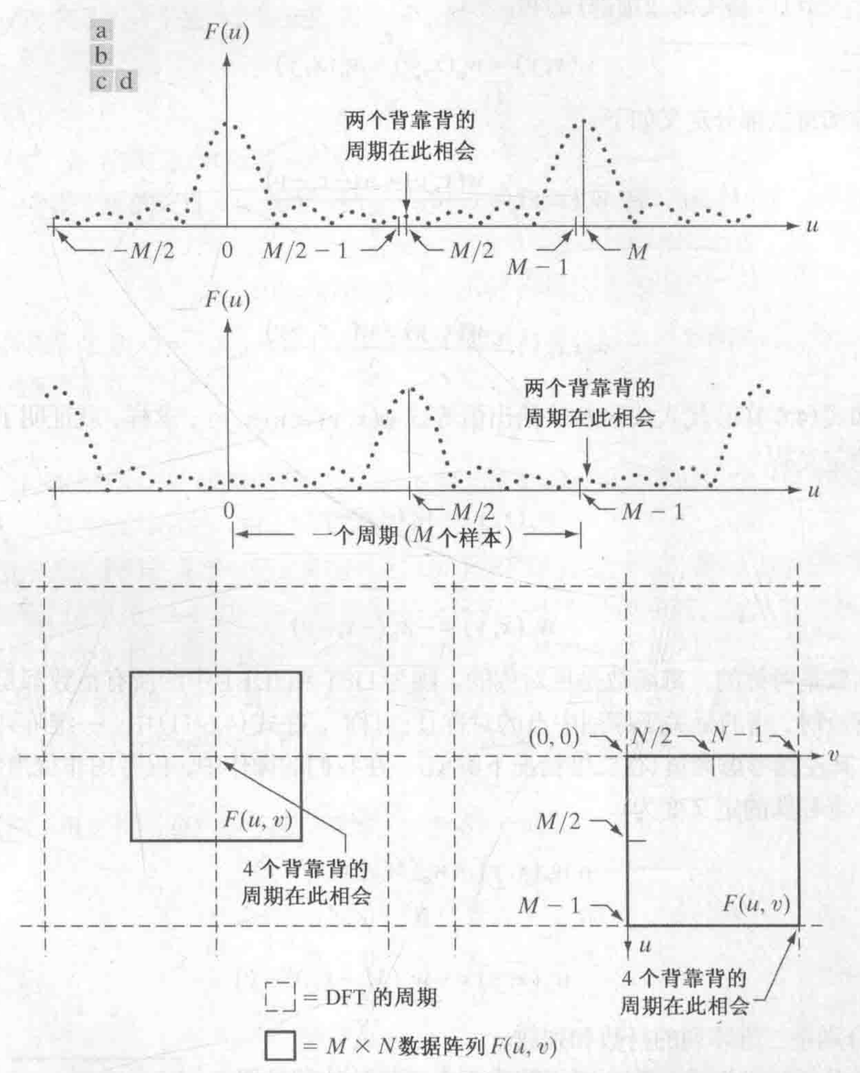


图 4.23　傅里叶变换的中心：(a) 显示了无限个周期的一个一维 DFT；(b) 在计算 $F(u)$ 之前，由 $(-1)^x$ 乘以 $f(x)$得到的移位后的 DFT；(c) 显示了无限个周期的一个二维 DFT。实线区域是由式(4.5-15)得到的 $M \times N$ 数据阵列 $F(u,v)$。该阵列由 4 个四分之一周期组成；(d) 在计算 $F(u,v)$ 之前，由 $(-1)^{x+y}$ 乘以 $f(x,y)$ 得到的移位后的 DFT。现在数据包含一个完整的位于中心的周期，这与图(b)一样

在二维情况下，用图形来显示这一情况有些困难，但原理相同，如图4.23(c)所示。代替使用两个半周期，现在有 4 个四分之一周期在点$(M/2, N/2)$处相会。虚线矩形对应于二维 DFT 的无限个周期。正像一维情况那样，若我们移动数据，以便 $F(0,0)$ 处在点 $(M/2, N/2)$ 处，则可简化其可视化程度。在式(4.6-3)中令 $(u_0, v_0) = (M/2, N/2)$，可得到如下表达式：

259
~
260

$$f(x,y)(-1)^{x+y} \Leftrightarrow F(u-M/2, v-N/2) \tag{4.6-8}$$

利用该式移动数据，以便如期望的那样，使$F(0,0)$处在由区间$[0,M-1]$和$[0,N-1]$定义的频率矩形的中心处。图 4.23(d)显示了这一结果。我们将在本章稍后的例 4.11 和图 4.24 中说明这些概念。

4.6.4 对称性

函数分析得到的一个重要结果是，任意实函数或虚函数$w(x,y)$均可表示为一个奇数部分和一个偶数部分(其中每个都可以是实部或虚部)的和：

$$w(x,y) = w_e(x,y) + w_o(x,y) \tag{4.6-9}$$

式中，偶数部分和奇数部分定义如下：

$$w_e(x,y) \triangleq \frac{w(x,y)+w(-x,-y)}{2} \tag{4.6-10a}$$

和

$$w_o(x,y) \triangleq \frac{w(x,y)-w(-x,-y)}{2} \tag{4.6-10b}$$

把式(4.6-10a)和式(4.6-10b)代入式(4.6-9)给出恒等式$w(x,y) \equiv w(x,y)$，这样，就证明了后一公式的正确性。由前面的定义有

$$w_e(x,y) = w_e(-x,-y) \tag{4.6-11a}$$

和

$$w_o(x,y) = -w_o(-x,-y) \tag{4.6-11b}$$

也就是说，偶函数是对称的，奇函数是反对称的。因为 DFT 和 IDFT 中的所有指数都是正的，当我们谈论对称(反对称)时，指的是关于序列中点的对称(反对称)。在式(4.6-11)中，一维阵列中心点右侧的指数考虑为正，其左侧考虑为负(在二维情况下类似)。在我们的操作中，仅考虑非负指数项更为方便，在这种情况下，奇和偶的定义变为

$$w_e(x,y) = w_e(M-x, N-y) \tag{4.6-12a}$$

和

$$w_o(x,y) = -w_o(M-x, N-y) \tag{4.6-12b}$$

式中，M和N分别是二维阵列的行数和列数。

> 若要自己动手证明一个奇函数的样本之和为零，可关于原点或跨越一个周期的任何其他区间画出一维正弦波的一个周期。

由初等数学分析我们知道，两个偶函数或两个奇函数的积是偶函数，而一个偶函数和一个奇函数的积是奇函数。另外，离散函数为奇函数的唯一方法是其所有样本之和为零。这些特性导致了一个重要结论，即对于任意一个离散偶函数w_e和一个离散奇函数w_o，有

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} w_e(x,y)w_o(x,y) = 0 \tag{4.6-13}$$

260 ~ 262

换句话说，由于式(4.6-13)的自变量是奇数，因此求和的结果为零。函数可以是实函数或复函数。

例 4.10　偶函数和奇函数。

虽然很容易就可看出连续函数的奇偶性，但在处理离散序列时，这些概念就不那么直观了。下面的说明将帮助我们弄清前面的概念。考虑一维序列

$$f = \{f(0) \quad f(1) \quad f(2) \quad f(3)\} = \{2 \quad 1 \quad 1 \quad 1\}$$

其中，$M=4$。为了检验偶性，必须满足条件 $f(x)=f(4-x)$，也就是说，我们要求

$$f(0)=f(4), f(2)=f(2), f(1)=f(3), f(3)=f(1)$$

因为 $f(4)$ 在被考察的范围之外，且它可以为任何值，所以 $f(0)$ 的值对于偶函数的测试不重要。我们看到，阵列中的值满足接下来的三个条件，则该序列是偶序列。事实上，我们得出结论，任何 4 点偶序列必须有下列形式：

$$\{a \;\; b \;\; c \;\; d\}$$

即在 4 点偶序列中，仅需第二点和最后一点相等。

奇序列有一个很有趣的特性，即其第一项 $w_o(0,0)$ 永远是 0，它遵循直接源自式(4.6-10b)的一个事实。考虑一维序列

$$g = \{g(0) \;\; g(1) \;\; g(2) \;\; g(3)\} = \{0 \;\; -1 \;\; 0 \;\; 1\}$$

由于序列中的各项满足 $g(x)=-g(4-x)$，因此我们很容易地确认这是一个奇序列。例如，$g(1)=-g(3)$。任何 4 点奇序列都有形式

$$\{0 \;\; -b \;\; 0 \;\; b\}$$

也就是说，当 M 是偶数时，一维奇序列有在位置 0 和 $M/2$ 处的点总是零的特性。当 M 是奇数时，第一项仍必须是零，其余项形成等值的一对，但符号相反。

前面的讨论指出，序列是偶序列还是奇序列还取决于序列的长度。例如，我们已经证明序列{0　−1　0　1}是奇序列。然而，序列{0　−1　0　1　0}既不是奇序列，也不是偶序列，虽然其“基本”结构表现为奇序列。在解释 DFT 的结果时，这是一个重要的问题。在本章稍后我们将说明偶函数和奇函数的 DFT 具有一些非常重要的特征。因此，了解一个函数是奇函数还是偶函数，在解释基于 DFT 的图像结果中扮演着重要的作用。

> 读者可自行证明这个二维序列是奇序列。

在二维情况中也有相同的基本考虑。例如，一个 6 × 6 的二维序列

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -2 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix}$$

是奇序列。但在添加另一行 0 和另一列 0 后，导致的序列就既不是奇序列也不是偶序列。注意，该阵列的内部结构是一个如 3.6.4 节讨论的 Sobel 模板。我们将在例 4.15 中再回顾这个模板。

利用前面的概念，我们可以建立 DFT 及其反变换的一些重要的对称性质。一个常用的性质是实函数 $f(x,y)$ 的傅里叶变换是共轭对称的，即

$$F^*(u,v) = F(-u,-v) \tag{4.6-14}$$

> 共轭对称也称为哈密特对称。术语反哈密特有时用于指共轭反对称。

如果 $f(x,y)$ 是虚函数，则其傅里叶变换是共轭反对称的：$F^*(-u,-v) = -F(u,v)$。式(4.6-14)的证明如下：

262
~
263

$$F^*(u,v)=\left[\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)}\right]^*=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f^*(x,y)\mathrm{e}^{\mathrm{j}2\pi(ux/M+vy/N)}$$

$$=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi([-u]x/M+[-v]y/N)}=F(-u,-v)$$

式中，第三步源自 $f(x,y)$ 是实函数这一事实。使用类似的方法可以证明虚函数的变换所表现出的共轭对称性。

表 4.1 列出了 DFT 的对称性和相关的性质，这些性质在数字图像处理中很有用。回忆可知，双箭头表示傅里叶变换对，即对于表中的任意一行，右边的性质可由拥有左边性质的函数的傅里叶变换来满足，反之亦然。例如，第 5 条性质可读为：实函数 $f(x, y)$的傅里叶变换是 $F^*(u,v)$，其中用$(-x, -y)$代替(x, y)，而 $f(x,y)$ 的离散傅里叶变换 $F(u, v)$ 是复函数，反之亦然。

表 4.1 二维 DFT 及其反变换的某些对称性质。$R(u, v)$ 和 $I(u,v)$分别是 $F(u, v)$ 的实部和虚部。术语复函数指出一个函数有非零实部和虚部

	空间域*		频率域*
1)	$f(x,y)$ 实函数	⇔	$F^*(u,v)=F(-u,-v)$
2)	$f(x,y)$ 虚函数	⇔	$F^*(-u,-v)=-F(u,v)$
3)	$f(x,y)$ 实函数	⇔	$R(u,v)$ 偶函数; $I(u,v)$ 奇函数
4)	$f(x,y)$ 虚函数	⇔	$R(u,v)$ 奇函数; $I(u,v)$ 偶函数
5)	$f(-x,-y)$ 实函数	⇔	$F^*(u,v)$ 复函数
6)	$f(-x,-y)$ 复函数	⇔	$F(-u,-v)$ 复函数
7)	$f^*(x,y)$ 复函数	⇔	$F^*(-u,-v)$ 复函数
8)	$f(x,y)$ 实函数和偶函数	⇔	$F(u,v)$ 实函数和偶函数
9)	$f(x,y)$ 实函数和奇函数	⇔	$F(u,v)$ 虚函数和奇函数
10)	$f(x,y)$ 虚函数和偶函数	⇔	$F(u,v)$ 虚函数和偶函数
11)	$f(x,y)$ 虚函数和奇函数	⇔	$F(u,v)$ 实函数和奇函数
12)	$f(x,y)$ 复函数和偶函数	⇔	$F(u,v)$ 复函数和偶函数
13)	$f(x,y)$ 复函数和奇函数	⇔	$F(u,v)$ 复函数和奇函数

*回忆可知，x, y, u 和 v 是离散(整数)变量，x 和 u 在区间 $[0, M-1]$ 内，y 和 v 在区间 $[0, N-1]$ 内。称一个复函数是偶函数意味着其实部和虚部都是偶函数，称一个复函数为奇函数同样意味着其实部和虚部都是奇函数。

例 4.11 表 4.1 中性质的一维说明。

根据早些时候讨论过的偶函数和奇函数的概念，以及例4.10 中的说明，下面的一维序列及它们的变换是表4.1 中所列性质的简单例子。右侧圆括号内的数字是 $F(u)$ 的单个元素；类似地，在最后两个特性中，对于 $f(x)$也类似。

性质	$f(x)$		$F(u)$
3	{1 2 3 4}	⇔	{(10)(−2 + 2j)(−2)(−2−2j)}
4	j{1 2 3 4}	⇔	{(2.5j)(0.5 − 0.5j)(− 0.5j)(− 0.5 − 0.5j)}
8	{2 1 1 1}	⇔	{(5)(1)(1)(1)}
9	{0 −1 0 1}	⇔	{(0)(2j)(0)(−2j)}
10	j{2 1 1 1}	⇔	{(5j)(j)(j)(j)}

263 ~ 265

11	j{0　−1　0　1}	⇔	{(0)(−2)(0)(2)}
12	{(4 + 4j)(3 + 2j)(0 + 2j)(3 + 2j)}	⇔	{(10 + 10j)(4 + 2j)(−2 + 2j)(4 + 2j)}
13	{(0 + 0j)(1 + 1j)(0 + 0j)(−1−1j)}	⇔	{(0 + 0j)(2−2j)(0 + 0j)(−2−2j)}

例如,在性质 3 中,我们看到有元素{1　2　3　4}的实函数具有傅里叶变换,该变换的实部{10　−2　−2　−2}是偶函数，虚部{0　2　0　−1}是奇函数。性质 8 告诉我们，一个实偶函数的傅里叶变换也是一个实偶函数。性质 12 说明，一个偶复函数的傅里叶变换也是复偶函数。其他性质的例子可以用同样的方法来分析。

例 4.12　证明表 4.1 中 DFT 的一些对称性质。

在这个例子中，我们证明表 4.1 中的一些性质，并熟练运用这些重要的性质，以便为解答本章末尾的一些习题打下基础。我们仅由左侧给出的性质证明右侧的性质。使用类似于此处给出的方式，可以用右侧给出的性质来证明左侧的性质。

考虑性质 3，它读做：如果 $f(x,y)$ 是实函数，则其 DFT 的实部是偶函数，而其虚部是奇函数；类似地，如果一个 DFT 分别具有偶函数的实部和奇函数的虚部，则其 IDFT 是一个实函数。这一性质的正式证明如下。$F(u,v)$ 通常是复函数，因此它可以表示为一个实部和一个虚部之和，即 $F(u,v)=R(u,v)+\mathrm{j}I(u,v)$。然后，有 $F^*(u,v)=R(u,v)-\mathrm{j}I(u,v)$。此外，有 $F(-u,-v)=R(-u,-v)+\mathrm{j}I(-u,-v)$。但是，正如早期证明的那样，如果 $f(x,y)$ 是实函数，则基于前两个等式有 $R(u,v)=R(-u,-v)$ 和 $I(u,v)=-I(-u,-v)$。根据式(4.6-11a)和式(4.6-11b)的观点，这就证明了 R 是一个偶函数，而 I 是一个奇函数。

下面，我们证明性质 8。如果 $f(x,y)$ 是一个实函数，那么由性质 3 可知 $F(u,v)$ 的实部是偶函数，因此为证明性质 8，我们所要做的所有工作是证明：如果 $f(x,y)$ 是实偶函数，则 $F(u,v)$ 的虚部是 0(即 F 是实函数)。证明步骤如下：

$$F(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)}$$

上式可写为

$$\begin{aligned}F(u,v)&=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f_r(x,y)]\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)}\\&=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f_r(x,y)]\mathrm{e}^{-\mathrm{j}2\pi(ux/M)}\mathrm{e}^{-\mathrm{j}2\pi(vy/N)}\\&=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\text{偶函数}][\text{偶函数}-\mathrm{j}\text{奇函数}][\text{偶函数}-\mathrm{j}\text{偶函数}]\\&=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\text{偶函数}]\text{偶函数}\cdot\text{偶函数}-2\mathrm{j}\text{偶函数}\cdot\text{偶函数}-\text{偶函数}\cdot\text{偶函数}\\&=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\text{偶函数}\cdot\text{偶函数}]-2\mathrm{j}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\text{偶函数}\cdot\text{偶函数}]-\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\text{偶函数}\cdot\text{偶函数}]\\&=\text{实函数}\end{aligned}$$

第 4 步来自欧拉公式及 cos 和 sin 分别是偶函数和奇函数的事实。由性质 8 我们还知道，除了是实函数之外，f 也是偶函数。在倒数第二行中，包含虚部分量的仅有一项是第二项，根据式(4.6-14)可知它是 0。这样，如果 f 是实偶函数，则 F 是实函数。正如早些时候解释的那样，因为 f 是实函数，所以 F 也是偶函数。结论得证。

最后，我们证明性质 6 的正确性。由 DFT 的定义，

$$\Im\{f(-x,-y)\}=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(-x,-y)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)}$$

265
~
266

由于周期性，有 $f(-x,-y)=f(M-x,N-y)$ 。如果现在定义 $m=M-x$ 和 $n=N-y$ ，则

$$\Im\{f(-x,-y)\}=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)\mathrm{e}^{-\mathrm{j}2\pi(u[M-m]/M+v[N-n]/N)}$$

(为确保求和本身是正确的，可用一个一维变换试一下，并手工扩展一些求和项。)因为 exp [-j2π(整数)]=1，它满足下式：

$$\Im\{f(-x,-y)\}=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)\mathrm{e}^{-\mathrm{j}2\pi(um/M+vn/N)}=F(-u,-v)$$

结论得证。

注意此处我们未更改变量。我们计算的是 $f(-x,-y)$ 的 DFT，因此只是简单地将该函数插入到了等式中，就像处理任何其他函数一样。

4.6.5 傅里叶谱和相角

因为二维 DFT 通常是复函数，因此可使用极坐标形式来表示：

$$F(u,v)=|F(u,v)|\mathrm{e}^{\mathrm{j}\phi(u,v)} \tag{4.6-15}$$

式中，幅度

$$|F(u,v)|=\left[R^2(u,v)+F^2(u,v)\right]^{1/2} \tag{4.6-16}$$

称为傅里叶谱(或频谱)，而

$$\phi(u,v)=\arctan\left[\frac{I(u,v)}{R(u,v)}\right] \tag{4.6-17}$$

称为相角。回忆4.2.1节的讨论，arctan必须使用一个四象限反正切来计算，如MATLAB的`atan2(Imag, Real)`函数。

最后，功率谱定义为

$$P(u,v)=|F(u,v)|^2=R^2(u,v)+I^2(u,v) \tag{4.6-18}$$

如之前那样，R 和 I 分别是 $F(u,v)$的实部和虚部，并且所有的计算直接对离散变量 $u=0,1,2,\cdots,M-1$ 和 $v=0,1,2,\cdots,N-1$ 进行。因此，$|F(u,v)|$，$\phi(u,v)$ 和 $P(u,v)$是大小为 $M\times N$ 的阵列。

实函数的傅里叶变换是共轭对称的［见式(4.6-14)］，这表明谱是关于原点偶对称的：

$$|F(u,v)|=|F(-u,-v)| \tag{4.6-19}$$

相角关于原点奇对称：

$$\phi(u,v)=-\phi(-u,-v) \tag{4.6-20}$$

它来自式(4.5-15)

$$F(0,0)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)$$

它指出零频率项与 $f(x,y)$的平均值成正比，即

$$F(0,0)=MN\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)=MN\bar{f}(x,y) \tag{4.6-21}$$



式中，$\bar{f}$ 表示 f 的平均值。从而有

$$|F(0,0)|=MN|\bar{f}(x,y)| \tag{4.6-22}$$

因为比例常数 MN 通常很大，典型地，$|F(0, 0)|$是频谱的最大成分，它可能比其他项大几个数量级。因为原点处的频率成分 u 和 v 是 0，所以 $F(0, 0)$有时称为变换的直流(dc)成分。这一术语源自电气工程，在电气工程中，"dc" 表示直流(即频率为零的电流)。

例 4.13 一个简单函数的二维傅里叶谱。

图4.24(a)显示了一幅简单的图像，图4.24(b)显示了该图像的谱，其值已被标定到区间[0, 255]，并以图像形式显示。空间域和频率域的原点都在左上角。在图4.24(b)中，两件事情很明显。如期望的那样，变换的原点的周围区域包含了最高值(在图像中表现为最亮)。但要注意，谱的 4 个角包含有类似的高值。原因是在前节中讨论的周期性。为中心化该谱，如式(4.6-8)指出的那样，在计算 DFT 之前我们简单地用$(-1)^{x+y}$乘以(a)中的图像。图 4.24(c)显示了结果，很明显它更易于观察(注意关于中心点的对称性)。由于直流项支配着谱的值，在显示的图像中，其他灰度的动态范围被压缩了。为了给出那些细节，如 3.2.2 节描述的那样，我们执行一次对数变换。图 4.24(d)给出了$(1 + \log|F(u, v)|)$的显示，细节再现的提升很明显。本章和后续几章中的多数谱均按这种方式标定过。

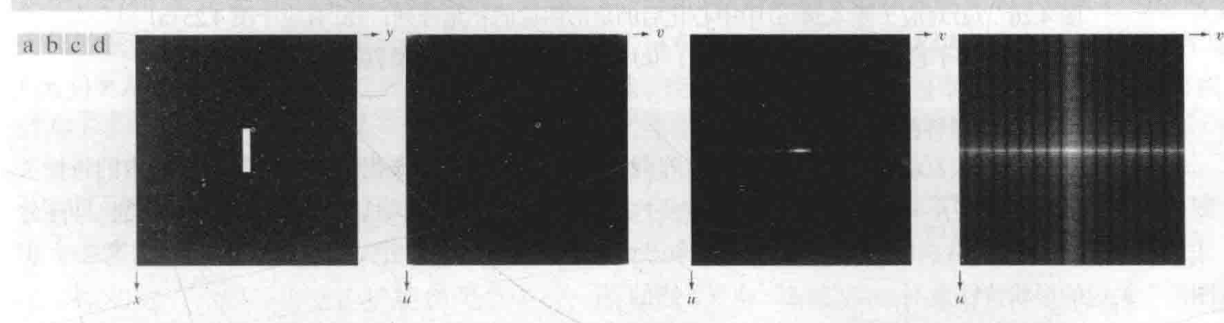


图 4.24 (a)图像；(b)在 4 个角显示亮点的谱；(c)中心化后的谱；(d)经对数变换后，增加了细节的结果。谱的零交叉在垂直方向较靠近，因为图(a)中的矩形在该方向上较长。全书所用的坐标习惯把空间域和频率域的原点放在左上角处

由式(4.6-4)和式(4.6-5)可知，谱对图像平移是不敏感的(指数项的绝对值是 1)，它随旋转图像以相同的角度旋转。图4.25说明了这些特性。图4.25(b)中的谱与图4.24(d)中的谱是相同的。很明显，图4.24(a)和图4.25(a)中的图像是不同的，因此，如果它们的傅里叶谱相同，则基于式(4.6-15)，它们的相角一定不同。图4.26 确认了这一点。图4.26(a)和(b)是图4.24(a)和图4.25(a)的 DFT 的相角阵列(以图像显示)。注意，相角图像之间缺少类似性，虽然它们所对应的图像间的唯一差别是简单的平移。通常，相角图像的视觉分析很少能得到直觉的信息。例如，在 45°角方向，我们的直觉是图4.26(a)中的相角应对应于图4.25(c)中的旋转图像，而不是图 4.24(a)的图像。事实上，如图 4.26(c)所示，旋转后图像的相角有一个很强的方向，它比 45°小得多。

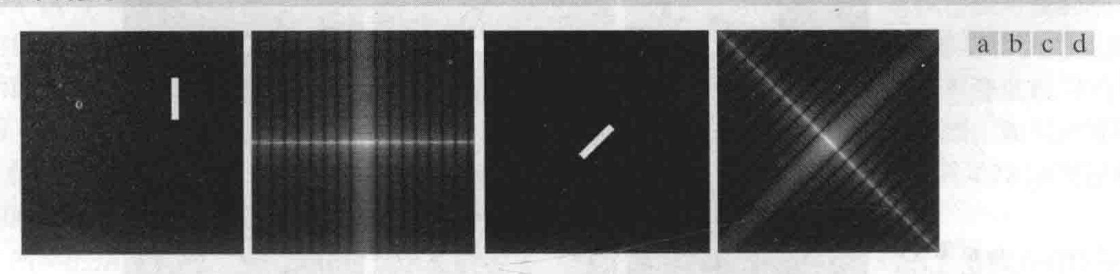


图 4.25 (a)平移后的图 4.24(a)中的矩形；(b)相应的谱；(c)旋转后的矩形；(d)相应的谱。对应于平移矩形的谱与图 4.24(a)中的原图像的谱相同

268 ~ 269

DFT 的谱的成发决定正弦波的幅度，这些正弦波结合起来可形成结果图像。在一幅图像的 DFT

中，在任何给定的频率处，较大的幅度意味着图像中该频率的正弦波比较突出。反之，较小的幅度意味着图像中出现的正弦波较少。如图4.26所示，虽然相角成分的贡献很不直观，但它恰恰很重要。相角是各个正弦成分关于原点的位移的度量。这样，当二维DFT的幅度是一个阵列时，其成分就决定了图像中的灰度，相应的相角则是一个角度阵列，它携带较多关于图像中可辨别物体定位的信息。下面的例子进一步阐明了这些概念。

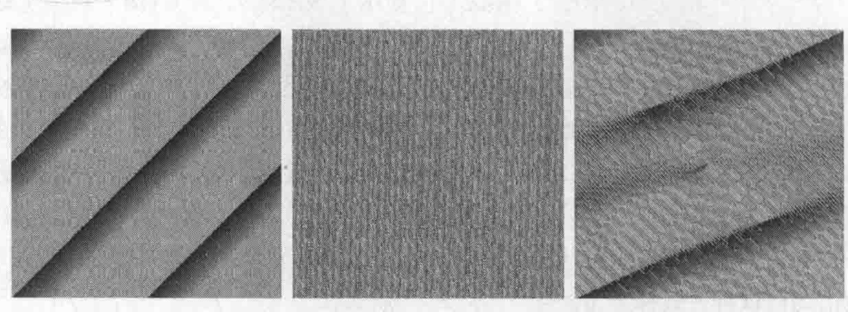


图 4.26 (a)对应于图4.24(a)中中心化后的矩形图像的相角阵列；(b)对应于图4.25(a)中平移后的图像的相角阵列；(c)对应于图4.25(c)中旋转后的图像的相角阵列

例 4.14 傅里叶谱和相角的进一步说明。

图4.27(b)是图4.27(a)的DFT的相角。该阵列将通过视觉分析引导我们将它与对应图像中的特性关联起来。然而，在决定形状特点时，相角的重要性在图4.27(c)中是很明显的，它仅用相角信息(即用公式 $|F(u, v)|=1$)计算式(4.6-15)的反傅里叶变换来得到。尽管灰度信息已丢失(记住，该信息由谱携带)，但图像中关键的形状特性是不会在图4.27(a)中弄错的。

图4.27(d)是仅使用式(4.6-15)的谱并计算其IDFT得到的。这意味着若将指数项置为1，这也同样意味着置相角为0，结果不会出现意外。它仅包含灰度信息，直流项占支配地位。图像中没有形状信息，因为相角已置为零。

最后，图4.27(e)和(f)再次证明了在确定一幅图像的特性内容时相角所起的支配作用。图4.27(e)是使用图4.24(a)中矩形的谱和对应于妇女图像的相角计算式(4.6-15)的IDFT得到的。妇女的形状显然支配了这一结果。反之，矩形支配了图4.27(f)，它是使用妇女图像的谱和矩形的相角计算得到的。

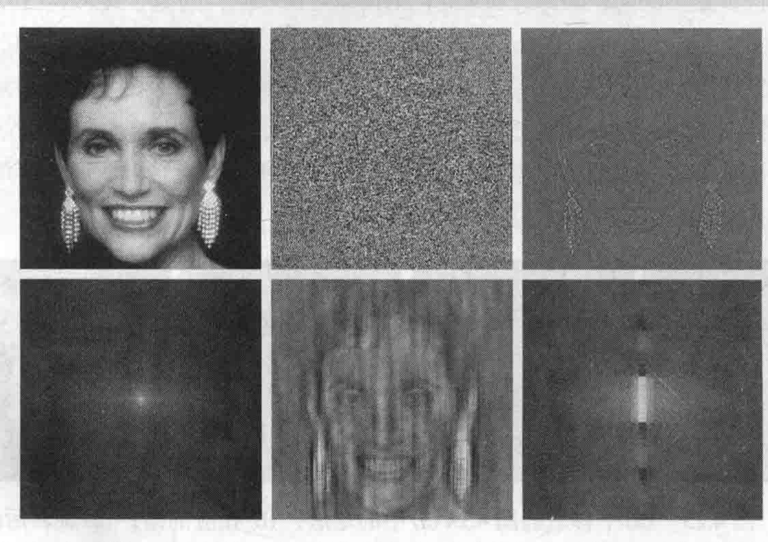


图 4.27 (a)妇女图像；(b)相角；(c)仅使用相角重建的妇女图像；(d)仅使用谱重建的妇女图像；(e)使用对应于妇女图像的相角和对应于图4.24(a)中矩形的谱重建的妇女图像；(f)使用矩形的相角和妇女图像的谱重建的图像

270 ~ 271

4.6.6　二维卷积定理

将式(4.4-10)扩展至两个变量可得到下面的二维循环卷积表达式：

$$f(x,y)\star h(x,y)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)h(x-m,y-n) \tag{4.6-23}$$

式中，$x=0,1,2,\cdots,M-1$，$y=0,1,2,\cdots,N-1$。如式(4.4-10)那样，式(4.6-23)给出了二维周期序列的一个周期。二维卷积定理由下面的表达式给出：

$$f(x,y)\star h(x,y)\Leftrightarrow F(u,v)H(u,v) \tag{4.6-24}$$

反之有

$$f(x,y)h(x,y)\Leftrightarrow F(u,v)\star H(u,v) \tag{4.6-25}$$

式中，F 和 H 是使用式(4.5-15)得到的，且与之前一样，双箭头用来表示表达式的左边和右边组成了傅里叶变换对。在本章其余部分，我们的兴趣是式(4.6-24)，该式表明乘积 $F(u,v)H(u,v)$ 的 IDFT 是 $f(x,y)\star h(x,y)$，即 f 和 h 的二维空间卷积。类似地，空间卷积的 DFT 是频率域中相应变换的乘积。式(4.6-24)是线性滤波的基础，且如 4.7 节中解释的那样，是本章讨论的所有滤波技术的基础。

> 我们将在 4.11 节中讨论计算 DFT 的有效方法。

因为在这里我们处理的是离散量，傅里叶变换的计算使用 DFT 算法执行。如果选择使用两个函数变换的乘积的 IDFT 来计算空间卷积，则必须考虑在 4.6.3 节中讨论过的周期性问题。下面我们给出一个关于该问题的一维的例子，然后将结论扩展到两个变量。图 4.28 左边一列执行两个函数 f 和 h 的卷积，它使用式(3.4-2)的一维等式，因为两个函数具有相同的尺寸，该卷积写为

$$f(x)\star h(x)=\sum_{m=0}^{399}f(x)h(x-m)$$

这个等式与式(4.4-10)相同，但对位移 x 的要求是它必须足够大，以便使反转的 h 完全滑过 f。换句话说，该过程包括：(1)关于原点求 h 的镜像(即使 h 旋转 180°)［见图4.28(c)］，(2)以数量 x 平移镜像函数［见图4.28(d)］，(3)对每个平移的 x 值，计算前面公式右边的全部乘积之和。依据图4.28，这意味着对每个 x 值用图4.28(d)中的函数乘以图4.28(a)的函数。位移 x 的范围是要求 h 完全滑过 f 的所有值。图4.28(c)显示了这两个函数的卷积。注意，卷积是位移变量 x 的函数，并且在这个例子中，要求 h 完全滑过 f 的 x 的范围是从 0 到 799。

如果使用 DFT 和卷积定理得到与图 4.28 左列相同的结果，就必须考虑 DFT 表达式中固有的周期。这与图 4.28(f)和图 4.28(g)中对两个周期函数进行卷积是等价的。卷积过程与我们刚才讨论的相同，但现在这两个函数是周期函数。按前一段介绍的方法处理这两个函数将得到图 4.28(j)所示的结果，很明显，它是不正确的。因为我们卷积两个周期函数，所以卷积本身也是周期的。在图 4.28 中，周期的靠近使它们互相干扰而导致所谓的缠绕错误。根据卷积定理，如果计算两个 400 点的函数 f 和 h 的 DFT，两个函数相乘，然后计算反 DFT，那么将得到如图 4.28(j)所示卷积的 400 点的错误一段。

所幸的是，解决缠绕错误问题很简单。考虑分别有 A 个样本和 B 个样本的两个函数 $f(x)$ 和 $h(x)$，可以证明(Brigham[1988])，如果把 0 添加到这两个函数中，使它们的长度相同，并用 P 来表示，那么可以按如下方式来选择，以避免缠绕：

$$P\geqslant A+B-1 \tag{4.6-26}$$

271 ~ 273

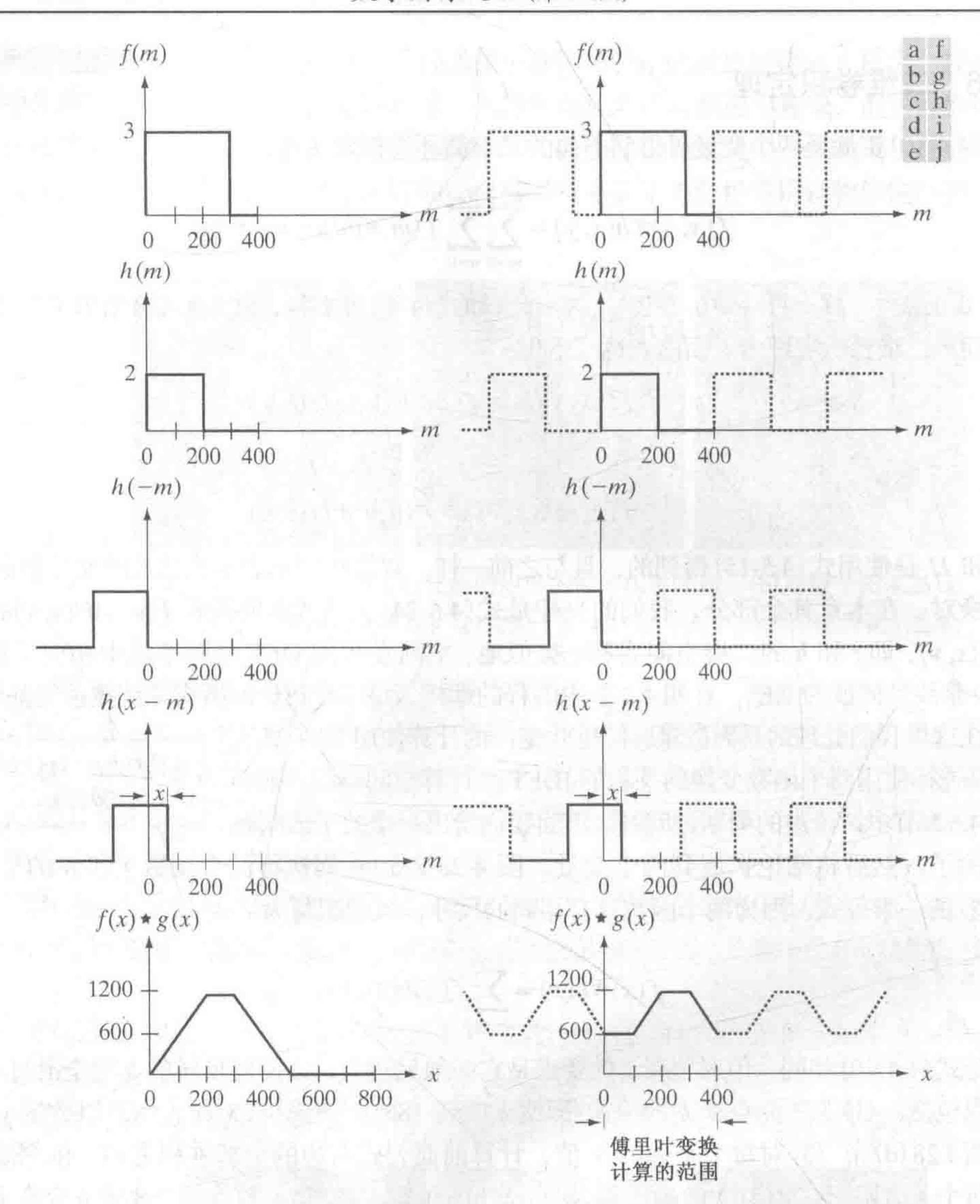


图 4.28　左列：用 3.4.2 节讨论的方法得到的两个离散函数的卷积。图(e)中的结果是正确的。右列：相同函数的卷积，但考虑了 DFT 的周期性。注意图(j)中来自邻近周期的数据是如何产生缠绕错误而得不到正确卷积结果的。要得到正确的结果，必须对函数进行填充

在我们的例子中，每个函数有 400 个点，我们应该使用的最小值是 $P = 799$，这意味着需要在每个函数的结尾处添加 399 个 0。这种处理称为 0 填充。作为练习，读者可以自己证明：如果图4.28(f)和图4.28(g)中的函数的周期已通过至少填充 399 个 0 的方法加长，那么结果就是一个周期卷积，其中每个周期都与图4.28(e)中的正确结果相等。对 DFT 使用卷积定理将得到与图4.28(e)相同的 799 点的空间函数。结论是：若要使第 3 章介绍的“直接”卷积公式法与 DFT 方法产生的卷积结果相同，后者必须对函数补足周期，再计算它们的变换。

> 零可添加到函数的开始处，或分别添加到函数的开始处和结尾处。但将零添加到函数的结尾处更为简单。

形象地表示二维情形下的例子较为困难，但我们可以得出关于缠绕错误及对函数补 0 的相同结论。令 $f(x, y)$ 和 $h(x, y)$ 分别是大小为 $A\times B$ 和 $C\times D$ 像素的图像阵列。循环卷积中的缠绕错误可以通过对这两个函数进行零填充来避免，方法如下：

273 ~ 274

$$f_p(x,y)=\begin{cases} f(x,y), & 0\leqslant x\leqslant A-1 \text{ 和 } 0\leqslant y\leqslant B-1 \\ 0, & A\leqslant x\leqslant P \text{ 或 } B\leqslant y\leqslant Q \end{cases} \tag{4.6-27}$$

和

$$h_p(x,y)\begin{cases} h(x,y), & 0\leqslant x\leqslant C-1 \text{ 和 } 0\leqslant y\leqslant D-1 \\ 0, & C\leqslant x\leqslant P \text{ 或 } D\leqslant y\leqslant Q \end{cases} \tag{4.6-28}$$

其中，

$$P\geqslant A+C-1 \tag{4.6-29}$$

和

$$Q\geqslant B+D-1 \tag{4.6-30}$$

填充后的图像大小为 $P\times Q$。如果两个阵列大小相同，都为 $M\times N$，则要求

$$P\geqslant 2M-1 \tag{4.6-31}$$

和

$$Q\geqslant 2N-1 \tag{4.6-32}$$

在 4.7.2 节，我们将给出一个例子来说明图像上出现缠绕错误所带来的影响。一般来说，DFT 算法对偶数尺寸的阵列执行较快，因此最好将 P 和 Q 选为满足上面方程的最小偶整数。如果两个阵列大小相同，则意味着 P 和 Q 选为该阵列大小的两倍。

在图4.28(a)和图4.28(b)中，两个函数的值在取样间隔的尾端可方便地变为 0。如果这两个函数中有一个函数或两个函数的值在取样区间的尾端不是 0，那么在把 0 添加到函数上以消除缠绕错误时，就将创建一个不连续的函数。这类似于用一个“盒子”与一个函数相乘，在频率域它意味着原始变换与一个 sinc 函数的卷积(见例 4.1)，依次地，这将造成一个由 sinc 函数的高频成分导致的所谓频率泄漏。泄漏会在图像上产生块效应。虽然漏泄从来未被完全消除过，但可以用另外一个函数乘以取样函数来明显地加以减少，这个函数在取样数据的两端平滑地过渡到接近于零，从而减弱“盒子”的(高频成分)急剧过渡。当在图像重建(如高清晰度图形)中希望保真度高时，这种称为开窗或切趾的方法应重点考虑。如果需要开窗，较好的方法是使用二维高斯函数(见4.8.3 节)。该函数的优点是其傅里叶变换也是高斯的，因此会产生较低的泄漏。

> 一个简单的切趾函数是位于取样数据中心的三角形，它会在取样数据的两端过渡到 0。这称为布特沃斯窗。其他一些常见的窗口是汉明窗和汉宁窗。我们甚至可以使用一个高斯函数。在 5.11.5 节中，我们将会回到加窗问题。

4.6.7　二维离散傅里叶变换性质的小结

表 4.2 总结了本章中介绍的主要的 DFT 定义。4.11.1 节将讨论可分性，4.11.2 节将讨论用正变换算法得到反变换。第 12 章将讨论相关。

表4.3 总结了一些重要的 DFT 对。虽然我们的关注点是离散函数，但表中的最后两项是仅针对连续变量推导的傅里叶变换对(注意连续变量符号的使用)。在此处包含这两项的原因是，它们在数字图像处理中非常有用。微分对可用于推导式(3.6-3)中定义的拉普拉斯变换的频率域公式(见问题 4.26)。高斯对将在 4.7.4 节讨论。

表 4.1 到表 4.3 总结了使用 DFT 时的有用性质。许多这类性质是本章剩余部分的关键内容，有些性质则会在后续几章中用到。

274 ~ 276

表 4.2 DFT 定义及相应表达式小结

名 称	表 达 式
1) $f(x, y)$的离散傅里叶变换(DFT)	$F(u,v)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)}$
2) $F(u, v)$的离散傅里叶反变换(IDFT)	$f(x,y)=\frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F(u,v)\mathrm{e}^{\mathrm{j}2\pi(ux/M+vy/N)}$
3) 极坐标表示	$F(u,v)=\|F(u,v)\|\mathrm{e}^{\mathrm{j}\phi(u,v)}$
4) 谱	$\|F(u,v)\|=\left[R^2(u,v)+I^2(u,v)\right]^{1/2}$ $R=\mathrm{Real}(F);\quad I=\mathrm{Imag}(F)$
5) 相角	$\phi(u,v)=\arctan\left[\frac{I(u,v)}{R(u,v)}\right]$
6) 功率谱	$P(u,v)=\|F(u,v)\|^2$
7) 均值	$\bar{f}(x,y)\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)=\frac{1}{MN}F(0,0)$
8) 周期性(k_1和k_2为整数)	$F(u,v)=F(u+k_1M,v)=F(u,v+k_2N)=F(u+k_1M,v+k_2N)$ $f(x,y)=f(x+k_1M,y)=f(x,y+k_2N)=f(x+k_1M,y+k_2N)$
9) 卷积	$f(x,y)\bigstar h(x,y)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f(m,n)h(x-m,y-n)$
10) 相关	$f(x,y)\star h(x,y)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}f^*(m,n)h(x+m,y+n)$
11) 可分性	二维 DFT 可以用先沿图像的行(列)计算一维 DFT 变换，然后沿结果的列(行)计算一维变换。见 4.11.1 节
12) 使用正变换算法得到傅里叶反变换	$MNf^*(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F^*(u,v)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)}$ 该式指出，将$F^*(u,v)$输入计算正变换的算法(上式右侧)中，将得到$MNf^*(x,y)$。取复共轭并除以MN就可给出希望的反变换。见 4.11.2 节

表 4.3 DFT 对的小结。第 12 项和第 13 项中的闭合表达式仅对连续变量有效。通过对闭合形式的连续表达式取样后，它们也可用于离散变量

名 称	DFT 对
1) 对称性	见表 4.1
2) 线性	$af_1(x,y)+bf_2(x,y)\Leftrightarrow aF_1(u,v)+bF_2(u,v)$
3) 平移性(一般)	$f(x,y)\mathrm{e}^{\mathrm{j}2\pi(u_0x/M+v_0y/N)}\Leftrightarrow F(u-u_0,v-v_0)$ $f(x-x_0,y-y_0)\Leftrightarrow F(u,v)\mathrm{e}^{-\mathrm{j}2\pi(ux_0/M+vy_0/N)}$
4) 平移到频率矩形的中心(M/2, N/2)	$f(x,y)(-1)^{x+y}\Leftrightarrow F(u-M/2,u-N/2)$ $f(x-M/2,y-N/2)\Leftrightarrow F(u,v)(-1)^{u+v}$
5) 旋转	$f(r,\theta+\theta_0)\Leftrightarrow F(\omega,\varphi+\theta_0)$ $x=r\cos\theta\ \ y=r\sin\theta\ \ u=\omega\cos\varphi\ \ v=\omega\sin\varphi$
6) 卷积定理*	$f(x,y)\bigstar h(x,y)\Leftrightarrow F(u,v)H(u,v)$ $f(x,y)h(x,y)\Leftrightarrow F(u,v)\bigstar H(u,v)$
7) 相关定理*	$f(x,y)\star h(x,y)\Leftrightarrow F^*(u,v)H(u,v)$ $f^*(x,y)h(x,y)\Leftrightarrow F(u,v)\star H(u,v)$
8) 离散单位冲激	$\delta(x,y)\Leftrightarrow 1$
9) 矩形函数	$\mathrm{rect}[a,b]\Leftrightarrow ab\frac{\sin(\pi ua)}{(\pi ua)}\frac{\sin(\pi ub)}{(\pi ub)}\mathrm{e}^{-\mathrm{j}\pi(ua+vb)}$
10) 正弦函数	$\sin(2\pi u_0x+2\pi v_0y)\Leftrightarrow \mathrm{j}\frac{1}{2}\left[\delta(u+Mu_0,v+Nv_0)-\delta(u-Mu_0,v-Nv_0)\right]$

276 ~ 277

续表

名　称	DFT 对
11）余弦函数	$\cos(2\pi u_0 x + 2\pi v_0 y) \Leftrightarrow \frac{1}{2}\left[\delta(u + Mu_0, v + Nu_0) + \delta(u - Mu_0, v - Nv_0)\right]$
下面的傅里叶变换对仅连续变量可推导，如之前那样，用 t 和 z 表示空间变量，用 μ 和 ν 表示频率变量。对连续函数取样后，这些结果可用于 DFT 处理。	
12）微分［右边的表达式假定 $f(\pm\infty,\pm\infty)=0$］	$\left(\frac{\partial}{\partial t}\right)^m\left(\frac{\partial}{\partial z}\right)^n f(t,z) \Leftrightarrow (\mathrm{j}2\pi\mu)^m(\mathrm{j}2\pi\nu)^n F(\mu,\nu)$ $\frac{\partial^m f(t,z)}{\partial t^m} \Leftrightarrow (\mathrm{j}2\pi\mu)^m F(\mu,\nu); \frac{\partial^n f(t,z)}{\partial z^n} \Leftrightarrow (\mathrm{j}2\pi\nu)^n F(\mu,\nu)$
13）高斯	$A2\pi\sigma^2 \mathrm{e}^{-2\pi^2\sigma^2(t^2+z^2)} \Leftrightarrow A\mathrm{e}^{-(\mu^2+\nu^2)/2\sigma^2}$（$A$ 是常数）

* 假定这些函数已经过 0 填充而得到了扩展。卷积和相关满足结合律、交换律和分配律。

4.7　频率域滤波基础

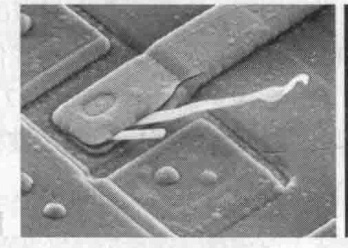
本节的内容是本章剩余部分讨论的所有滤波技术的基础。

4.7.1　频率域的其他特性

我们从观察式(4.5-15)开始，式中 $F(u, v)$的每一项都包含有用指数项修改过的 $f(x, y)$的所有值。这样，除了不太重要的情况外，将一幅图像的某些特定成分与其变换直接联系起来通常是不可能的，然而，关于傅里叶变换的频率成分和一幅图像的空间特性间的关系可以做出某些一般的叙述。例如，因为频率直接关系到空间变化率，因此直观地将傅里叶变换中的频率与图像中的亮度变化模式联系起来并不困难。我们在 4.6.5 节已经说过，变化最慢的频率成分($u = v = 0$)与图像的平均灰度成正比。当我们远离变换的原点时，低频对应于图像中变化缓慢的灰度成分。例如，在一幅房间图像中，这些可能对应于墙壁和地板的平滑灰度变化。当我们从原点移开更远一些时，较高的频率开始对应于图像中越来越快的灰度变化。这些变化是物体的边缘，或由灰度急剧变化来表征的其他图像成分。

如 2.6.7 节介绍的那样，频率域中的滤波技术是以如下处理为基础的：修改傅里叶变换以达到特殊目的，然后计算 IDFT 返回到图像域。遵循式(4.6-15)，我们所用变换的两个分量是变换的幅度(谱)和相角。4.6.5 节涵盖了变换的这两个分量的基本性质。在那里，我们已了解到，相角分量的视觉分析通常并不是很有用，然而，产生该谱的图像的大致特点提供了某些很有用的指导。例如，考虑图 4.29(a)，它是一块集成电路的扫描电子显微图像，放大了近 2500 倍。除了器件本身的有趣结构之外，我们注意到两个主要特征：近似 ± 45°方向的很强边缘和两个由热感应故障导致的白色氧化突起物。图 4.29(b)中的傅里叶谱显示了沿 ± 45°方向的突出成分，这些成分对应于刚刚提到的边缘。沿垂直轴仔细观察，可看到一个稍微左偏垂直轴的垂直成分，该成分是由氧化突起物的边缘导致的。注意该频率成分与垂直轴的夹角是如何对应于白色长物体(关于水平轴的)的斜度的，还要注意垂直频率成分中的零是如何对应于氧化突起物的狭窄垂直范围的。

a b



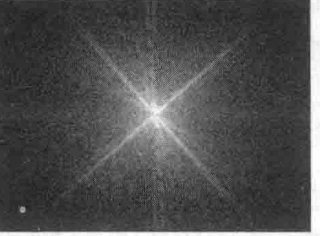


图 4.29　(a)已损坏集成电路的 SEM 图像；(b)图(a)的傅里叶谱(原图像由加拿大安大略省哈密尔顿市 McMaster 大学材料研究所的 J. M. Hudak 博士提供)

277 ~ 278

这些对应关系通常是在频率域与空间域之间建立的典型联系。正如本章稍后将要说明的那样，恰恰是这些总体联系的类型与前面提到的频率内容和图像灰度变化率之间的伴随关系，导致了某些非常有用的结果。下一节将说明修改图4.29(a)的变换中的各种频率范围的效果。

4.7.2 频率域滤波基础

频率域滤波是指先修改一幅图像的傅里叶变换，然后计算其反变换，得到处理后的结果。由此，若给定一幅大小为 $M \times N$ 的数字图像 $f(x, y)$，则我们感兴趣的基本滤波公式有如下形式：

$$g(x,y) = \Im^{-1}\left[H(u,v)F(u,v)\right] \tag{4.7-1}$$

式中，$\Im^{-1}$ 是 IDFT，$F(u, v)$是输入图像 $f(x, y)$的 DFT，$H(u, v)$是滤波函数(简称为滤波器或滤波传递函数)，$g(x, y)$是滤波后的(输出)图像。函数 F, H 和 g 是大小与输入图像相同的 $M \times N$ 阵列。如 2.6.1 节定义的那样，乘积 $H(u, v)F(u, v)$是由阵列相乘形成的。滤波函数修改输入图像的变换来得到处理后的输出 $g(x, y)$。使用关于中心对称的函数可明显简化 $H(u, v)$的技术条件，它要求 $F(u, v)$也被中心化。如 4.6.3 节解释的那样，这是通过在计算变换前使用 $(-1)^{x+y}$ 乘以输入图像来完成的[1]。

> 如果 H 是实对称函数而 f 是实函数(通常如此)，则式(4.7-1)中的 IDFT 理论上应生成实数量。实际上，该反变换通常包含有由舍入误差和其他计算错误引起的寄生复数项。因此，我们通常取 IDFT 的实部来形成函数 g。

现在，我们已能考虑滤波过程的一些细节。我们能构建的最简单的滤波器之一是滤波器 $H(u, v)$，它在变换的中心位置是 0，而在其他位置是 1。当我们建立乘积 $H(u, v)F(u, v)$时，该滤波器将抑制 $F(u, v)$的直流项而“通过”(即保持不变)所有其他项。由式(4.6-21)可知，直流项决定图像的平均灰度，因此将其置 0 会把输出图像的平均灰度减小为 0。图4.30 显示了使用式(4.7-1)时这一操作的结果。如期望的那样，图像变得更暗［零平均表明存在负灰度。因此，尽管图4.30 说明了原理，但不是原图像的真实描述，因为所有负灰度为显示目的都被修剪过(置为0)］。

如此前说明的那样，变换中的低频与图像中缓慢变化的灰度成分有关，如室内的墙壁和室外少云的天空等。另一方面，高频由灰度的尖锐过渡造成，如边缘和噪声等。因此，我们认为衰减高频而通过低频的滤波器 $H(u, v)$(近似称为低通滤波器)将模糊一幅图像，具有相反特性的滤波器(称为高通滤波器)将增强尖锐的细节，但会降低图像的对比度。图4.31 说明了这些效应。注意图4.31(e)和图4.30 之间的相似性。原因是所示的高通滤波器消去了直流项，导致了与图4.30 相同的基本效果。对滤波器加上一个小常数不会影响尖锐性，但它的确能防止直流项的消除，并保留色调，如图4.31(f)所示。

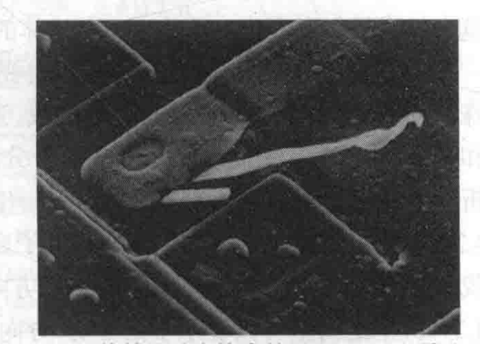


图 4.30　将傅里叶变换中的 $F(M/2, N/2)$项置 0，对图 4.29(a)中的图像滤波后的结果

式(4.7-1)涉及频率域中两个函数的乘积，由卷积定理，这意味着空间域的卷积。我们从 4.6.6 节的讨论知道，如果问题中的函数未被填充，可以料到会产生混淆误差。当我们用式(4.7-1)而没有填充时，考虑一下会发生什么情况。图 4.32(a)显示了一幅简单的图像，图 4.32(b)是使用图 4.31(a)所示形式的高斯低通滤波器对图像低通滤波后的结果。如预料的那样，图像被模糊，但模糊不均匀；顶

279 ~ 280

① 许多二维 DFT 的软件实现(如 MATLAB)并不要求变换关于中心对称。这意味着我们必须安排滤波函数，以便其数据格式与非中心对称变换的数据格式相同(即让原点在左上角)。最终结果是生成和显示滤波器更为困难。在我们的讨论中，为便于显示，我们使用中心对称的变换，这对于清楚理解滤波的概念至关重要。只要保持了一致性，实践中可使用任意一种方法。

部的白边被模糊，而两侧的白边未被模糊。在应用式(4.7-1)之前，根据式(4.6-31)和式(4.6-32)填充输入图像，将导致图4.32(c)所示的滤波后的图像。这正是我们期望的结果。

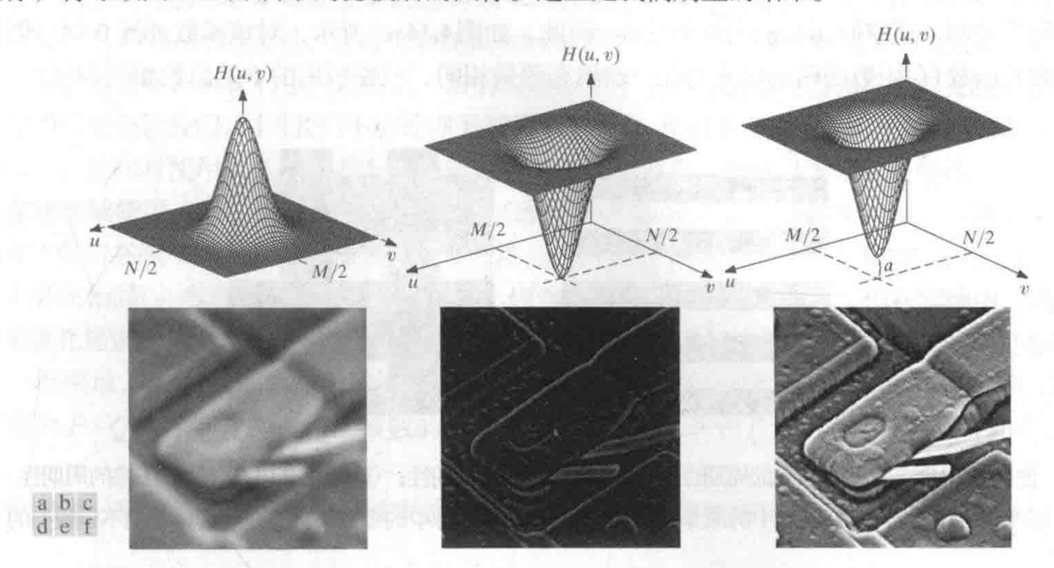


图 4.31　上一行：频率域滤波器。下一行：使用式(4.7-1)得到的相应的滤波后的图像。在图(c)中我们使用 $a = 0.85$ 来得到图(f)(滤波器本身的高度是 1)。请将图(f)与图 4.29(a)比较

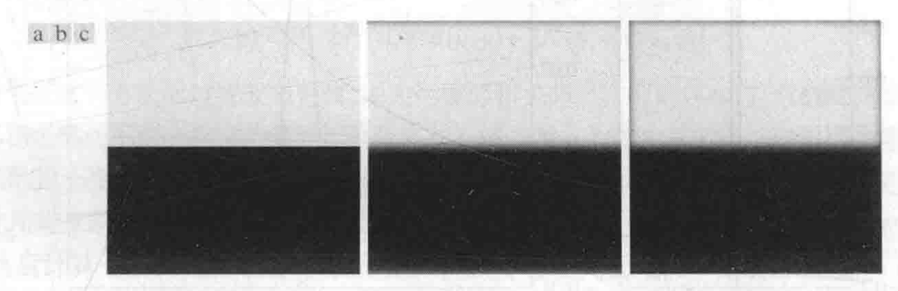


图 4.32　(a)一幅简单的图像；(b)无填充时使用高斯低通滤波器的模糊结果；(c)填充后的低通滤波的结果。请比较图(b)和图(c)中垂直边缘的明亮区域

图4.33 给出了图4.32(b)和(c)之间差异的原因。图 4.33 中的虚线区域对应图 4.32(a)中的图像。如 4.6.3 节解释的那样，图4.33(a)显示了使用 DFT 的固有周期性。想象一下该模糊滤波器的空间表示与该图像卷积。当滤波器通过虚线图像区域的上方时，它将包围该图像的一部分及其右上方周期图像底部的一部分。当一个暗区域和一个亮区域在滤波器下方时，结果是一幅中等灰度的模糊图像。然而，当滤波器通过图像的右上侧时，滤波器将只包围图像中的亮区域及其右侧的相邻区域。一个常数的平均仍是该常数，因此在这个区域滤波将无效，图 4.32(b)给出了结果。用 0 填充图像后，在周期序列的周围建立了一个平坦的边界，如图 4.33(b)所示。将模糊函数与图 4.33(b)所示填充过的“马赛克”图像卷积，给出了图 4.33(c)所示的正确结果。从这个例子可以看到，错误地填充一幅图像会导致错误的结果。如果滤波的目的仅是粗糙的视觉分析，那么有时可以跳过填充步骤。

到目前为止，讨论集中在输入图像的填充上，但还包含一个滤波器，该滤波器可以在空间域或频率域中来规定。然而，填充是在空间域完成的，这就提出了空间填充和直接在频率域指定滤波器间的关系这个重要问题。

乍看之下，一种可能的结论是：处理频率域滤波器填充的一种方法是，构建一个与图像尺寸相同的滤波器，计算该滤波器的 IDFT 得到相应的空间滤波器，在空间域填充这个空间滤波器，然后计算其 DFT 返回到频率域。图4.34所示的一维例子说明了这种方法的缺陷。图4.34(a)显示了频率域中的一

281 ~ 282

个一维理想低通滤波器。该滤波器是实的和偶对称的，因此由表 4.1 中的性质 8 可知其 IDFT 也是实的和偶对称的。图4.34(b)显示了用$(-1)^u$乘以频率域滤波器的元素的结果，并计算了其 IDFT 以得到相应的空间滤波器。该空间函数的两端不是 0，因此，如图4.34(c)所示，对该函数进行 0 填充创建了两个不连续的函数(在函数的两端填充与在一端填充效果相同，只要所用的 0 的总数相同即可)。

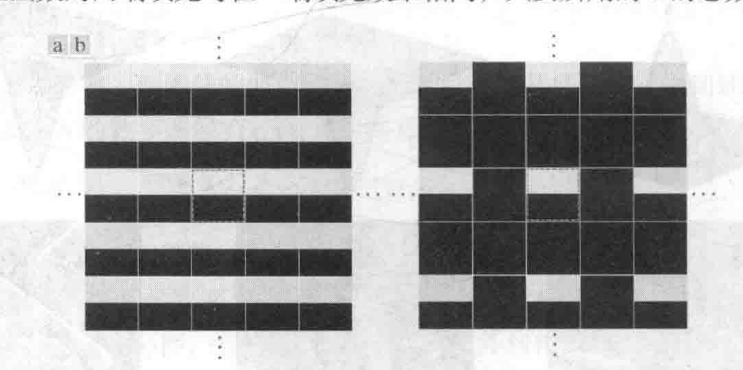


图 4.33 使用DFT时二维图像固有的周期性。(a)未填充图像的周期性；(b)使用0填充(黑)后图像的周期性。中心的虚线区域对应于图 4.32(a)中的图像(为清楚起见，两幅图像中的细白线重叠在一起，它们不是数据的一部分)

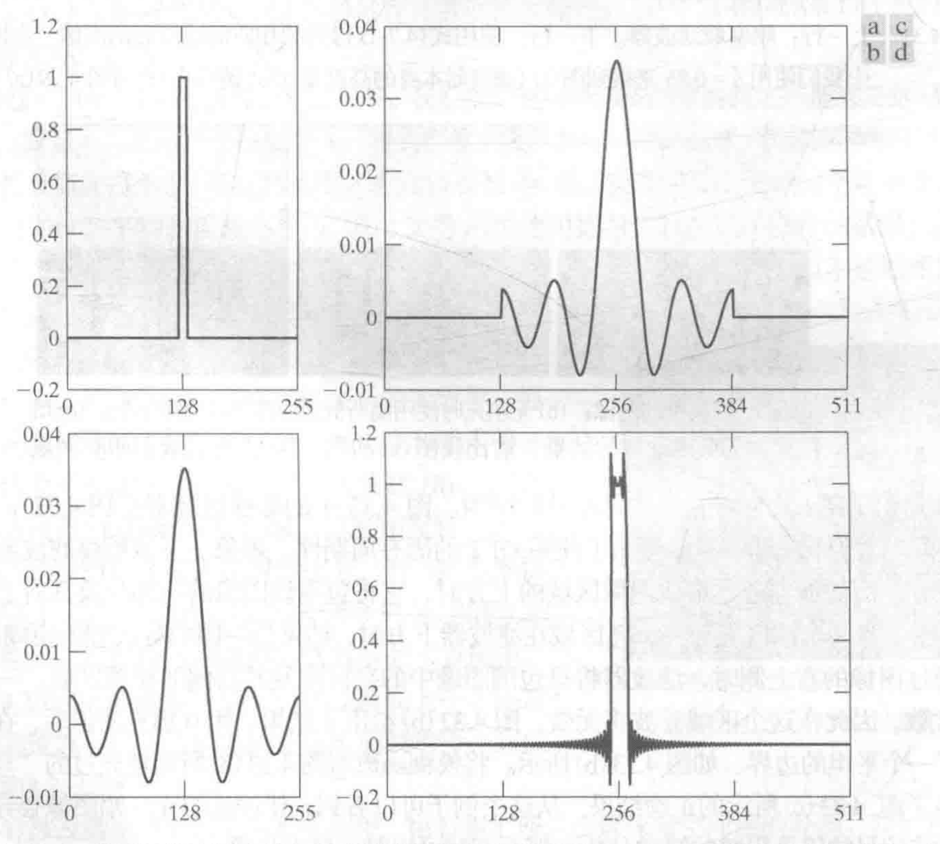


图 4.34 (a)在(中心对称的)频率域中指定的原始滤波器；(b)计算图(a)的 IDFT 得到的空间表示；(c)将图(b)填充至其两倍长度后的结果(注意不连续点)；(d)通过计算图(c)的 DFT 得到的相应的频率域滤波器。注意，图(c)中的不连续导致了振铃现象(为简化视觉分析，点已连接起来，因此曲线看起来是连续的)

为了回到频率域，我们计算填充后的空间滤波器的 DFT。图4.34(d)显示了这一结果。正如在例 4.1 中预料的那样，空间滤波器中的不连续在其频率域中产生了振铃。如果以另一种方法来观察，我们从一个例子知道，一个盒状函数的傅里叶

理想滤波器的定义见4.3.3 节的末尾。

282 ~ 283

变换是一个其频率成分无限扩展的 sinc 函数，因此我们期望合状函数的反变换也有相同的属性。也就是说，一个理想的(盒状)频率域滤波器的空间表示具有无限扩展的成分。因此，任何实现 0 填充的滤波器的空间截断将引入不连续性，它通常会在频率域导致振铃现象(在这种情况下，如果它在 0 交叉处做填充，可避免截尾，但我们感兴趣的是一般的过程，而不是所有的滤波器都有 0 交叉的过程)。

前面的结果告诉我们，因为我们不能处理无限数量的成分，所以不可能使用理想的频率域滤波器[见图 4.34(a)]，并同时使用零填充来避免缠绕错误。因此，需要有一个可以接受的限制性决策。我们的目标是在频率域使用一个特定形状的滤波器(包括理想滤波器)而不必涉及截断问题。一种方法是对图像 0 填充，然后在频率域创建滤波器，其尺寸与填充过的图像一样（记住，使用 DFT 时，图像和滤波器的大小必须相同)。当然，这将导致缠绕误差，因为未对该滤波器进行填充。但在实践中，该错误可通过图像填充提供的间隔有效地减轻，并尽力消除振铃现象。平滑滤波器(如图4.31中的那些)存在的问题更少。特别地，为了直接在频率域使用特定形状的滤波器进行处理，本章中我们采用的方法是把图像填充为 $P\times Q$ 大小，并构造相同维数的滤波器。如前所述，P 和 Q 分别由式(4.6-29)和式(4.6-30)给出。

我们将以分析滤波变换的相角来结束这一节。因为 DFT 是复数阵列，我们可以将它表示为实部和虚部：

$$F(u,v)=R(u,v)+\mathrm{j}I(u,v) \tag{4.7-2}$$

然后，式(4.7-1)变为

$$g(x,y)=\Im^{-1}\left[H(u,v)R(u,v)+\mathrm{j}H(u,v)I(u,v)\right] \tag{4.7-3}$$

相角不能以刚刚描述的方法通过滤波来更改，因为当虚部和实部之比由式(4.6-17)形成时，就消去了 $H(u,v)$。等同影响实部和虚部而不影响相位的滤波器称为零相移滤波器。本章中仅考虑这种类型的滤波器。

即使相角的很小变化，也会对滤波输出有很大的影响(通常是我们不希望的)。图4.35说明了如标量一样简单的变化所带来的影响。图4.35(a)显示了用 0.5 乘以式(4.6-15)中的角度阵列所得到的图像，它不改变 $|F(u,v)|$，然后计算 IDFT。基本形状保持不变，但灰度分布失真很大。图4.35(b)显示了用 0.25 乘以相角的结果。该图像几乎无法识别。

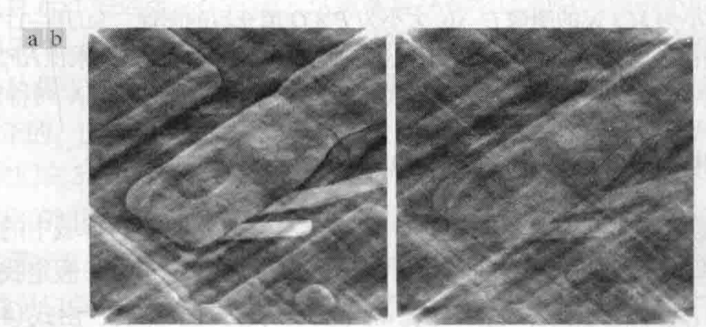


图 4.35　(a)使用 0.5 乘以式(4.6-15)中的相角然后计算 IDFT 得到的图像；(b)使用 0.25 乘以相角得到的结果。两种情况下的谱没有变化

4.7.3　频率域滤波步骤小结

前两节的内容可以总结如下：

1. 给定一幅大小为 $M\times N$ 的输入图像 $f(x,y)$，从式(4.6-31)和式(4.6-32)得到填充参数 P 和 Q。通常选择 $P=2M$ 和 $Q=2N$。
2. 对 $f(x,y)$添加必要数量的 0，形成大小为 $P\times Q$ 的填充图像 $f_p(x,y)$。

283～285

3. 用$(-1)^{x+y}$乘以$f_p(x, y)$，移到其变换的中心。
4. 计算来自步骤 3 的图像的 DFT，得到$F(u, v)$。
5. 生成一个实的、对称的滤波函数$H(u, v)$，其大小为$P \times Q$，中心在$(P/2, Q/2)$处[①]。用阵列相乘形成乘积$G(u, v) = H(u, v)F(u, v)$；即$G(i, k) = H(i, k)F(i, k)$。
6. 得到处理后的图像：

> 如前所述，关于中心对称有助于形象地描述滤波过程并生成滤波函数本身，但它不是基本的需求。

$$g_p(x, y) = \left\{ \text{real}\left[\Im^{-1}\left[G(u, v) \right] \right] \right\} (-1)^{x+y}$$

其中，为忽略由于计算不准确导致的寄生复成分，选择了实部，下标p指出我们处理的是填充后的阵列。

7. 从$g_p(x, y)$的左上象限提取$M \times N$区域，得到最终处理结果$g(x, y)$。

图4.36说明了上述步骤。图中的符号解释了每个图像源。如果放大，图 4.36(c)将会显示图像中插入的黑点，因为为了显示，负灰度被裁剪为 0。注意图4.36(h)中图像的暗边，它是使用零填充处理后由低通滤波造成的。

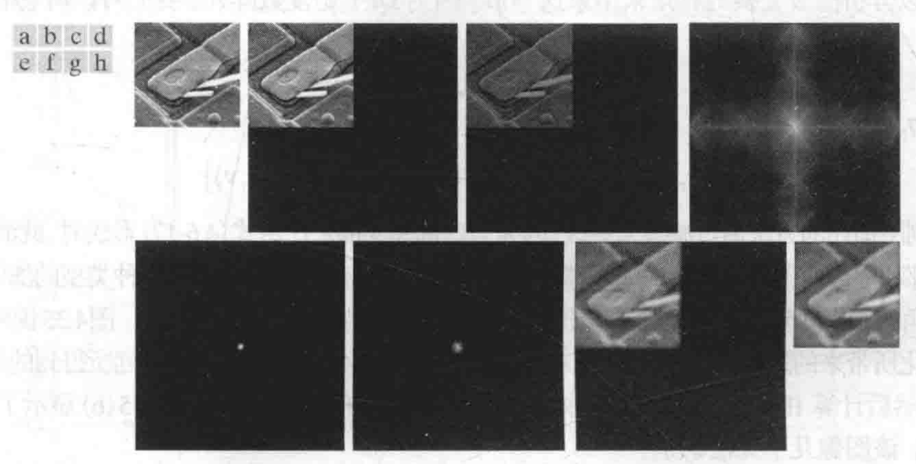


图 4.36　(a)一幅大小为$M \times N$的图像f；(b)大小为$P \times Q$填充后的图像；(c)用$(-1)^{x+y}$乘以f_p的结果；(d)F_p的谱；(e)大小为$P \times Q$的关于中心对称的高斯低通滤波器；(f)乘积HF_p的谱；(g)$(-1)^{x+y}$与HF_p的 IDFT 的实部的乘积g_p；(h)裁剪g_p的前M行和前N列得到的最终结果g

4.7.4　空间和频率域滤波间的对应

空间域和频率域滤波间的纽带是卷积定理。在 4.7.2 节，我们把频率域中的滤波定义为滤波函数$H(u, v)$与输入图像的傅里叶变换$F(u, v)$的乘积。给定一个滤波器$H(u, v)$，假定我们想要找出其空间域的等价表示。如果令$f(x, y) = \delta(x, y)$，它来自表 4.3 的$F(u, v) = 1$。然后，由式(4.7-1)可知滤波后的输出是$\Im^{-1}\left[H(u,v) \right]$。但这是频率域滤波器的反变换，它对应空间域的滤波器。反过来，遵循类似的分析和卷积定理，给定一个空间滤波器，我们可以用该空间滤波器的傅里叶正变换得到其频率域表示。因此，两个滤波器形成了傅里叶变换对：

$$h(x, y) \Leftrightarrow H(u, v) \tag{4.7-4}$$

式中，$h(x, y)$是一个空间滤波器。因为该滤波器可以由频率域滤波器对一个冲激的响应得到，所以$h(x, y)$

285 ~ 286

① 如果$H(u, v)$由一个给定的空间滤波器$h(x, y)$生成，则可按如下方式形成$h_p(x, y)$：将该空间滤波器的大小填充为$P \times Q$，使用$(-1)^{x+y}$乘以扩展后的阵列，并计算结果的 DFT 来得到$H(u, v)$。例 4.15 给出了这个过程。

有时称为 $H(u, v)$的脉冲响应。此外，由于式(4.7-4)的离散实现中的所有数值都是有限的，这样的滤波器称为有限冲激响应(FIR)滤波器。这些仅是在本书中考虑的一类线性空间滤波器。

我们在3.4.1节中介绍了空间卷积，并讨论了其与式(3.4-2)有关的实现，它涉及不同尺寸的卷积函数。当我们根据卷积定理和 DFT 谈到空间卷积时，正如图 4.28 解释的那样，这意味着我们对周期函数执行卷积操作。由于这一原因，如之前解释的那样，式(4.6-23)称为循环卷积。此外，在 DFT 的意义上，卷积包含相同尺寸的函数，反之，在式(3.4-2)中，函数的尺寸通常不同。

在实践中，我们更愿使用式(3.4-2)和较小的滤波器模板来实现卷积滤波，因为它在硬件和/或固件实现时速度更快，且更容易。然而，滤波的概念在频率域更直观。取两个域中特性的优点的一种方法是，在频率域规定一个滤波器，计算它的 IDFT，然后如构建较小的空间滤波模板那样，用反变换的结果得到全尺寸的空间滤波器（更正式的方法将在4.11.4节中介绍)。这将在下一节中再度说明。在这一节稍后，我们将给出相反的过程，这里给定一个小空间滤波器，我们得到其全尺寸频率域表示。这种方法在频率域分析较小的空间滤波的特性时是很有用的。记住，在下面的讨论中，傅里叶变换及其反变换是线性处理(见习题 4.14)，因此讨论只限于线性滤波。

在下面的讨论中，我们使用高斯滤波器来说明如何运用频率域滤波器指定第 3 章中讨论的一些小模板的系数。基于高斯函数的滤波器特别重要，因为如表 4.3 注释的那样，一个高斯函数的正、反傅里叶变换都是实高斯函数。为便于说明原理，这里仅讨论一维高斯滤波器。二维高斯滤波器在这一章稍后再讨论。

令 $H(u)$表示一维频率域高斯滤波器：

$$H(u) = A\mathrm{e}^{-u^2/2\sigma^2} \tag{4.7-5}$$

式中，σ 是高斯曲线的标准差。空间域中的相应滤波器由 $H(u)$的傅里叶反变换得到(见习题 4.31)：

$$h(x) = \sqrt{2\pi}\sigma A\mathrm{e}^{-2\pi^2\sigma^2x^2} \tag{4.7-6}$$

这些公式[①]很重要，原因有二：(1)它们是一个傅里叶变换对，两个成分都是高斯函数和实函数。这便于分析，因为我们不必涉及复数。另外，高斯曲线很直观，并易于操作。(2)函数表现为互易的。当 $H(u)$ 有一个较宽的外形时(σ 值较大)，$h(x)$有较窄的外形，反之亦然。事实上，当 σ 接近于无限时，$H(u)$ 接近一个常数函数，并且 $h(x)$ 趋近于一个冲激，这意味着在频率域和空间域都不能滤波。

图4.37(a)和图4.37(b)显示了频率域高斯低通滤波器和空间域对应低通滤波器的曲线。假定我们想要使用图4.37(b)中 $h(x)$的曲线形状作为指导来规定一个小空间模板的系数。这两个滤波器之间的关键相似性是它们的所有值都是正的。这样，我们就得出一个结论：我们使用一个全部带正系数的模板就可在空间域中实现低通滤波(如我们在 3.5.1 节中所做的那样)。作为参考，图4.37(b)中显示了在 3.5.1 节中讨论的两个模板。注意前一段讨论的滤波器宽度间的互易关系。频率域滤波器越窄，其衰减的低频越多，引起的模糊越大。在空间域，这意味着必须使用较大的模板来增加模糊，正如例 3.13 中说明的那样。

更复杂的滤波器可以使用式(4.7-5)给出的基本高斯函数来构建。例如，我们可以用高斯函数的差来构造一个高通滤波器：

$$H(u) = A\mathrm{e}^{-u^2/2\sigma_1^2} - B\mathrm{e}^{-u^2/2\sigma_2^2} \tag{4.7-7}$$

式中，$A \geqslant B, \sigma_1 > \sigma_2$。空间域中的相应滤波器是

$$h(x) = \sqrt{2\pi}\sigma_1 A\mathrm{e}^{-2\pi^2\sigma_1^2x^2} - \sqrt{2\pi}\sigma_2 B\mathrm{e}^{-2\pi^2\sigma_2^2x^2} \tag{4.7-8}$$

图4.37(c)和图4.37(d)显示了这两个公式的曲线。我们再次注意到宽度的互易性，但这里最重要的特性

286 ~ 288

① 像表 4.3 中提及的那样，高斯函数的闭合形式的正反傅里叶变换仅对连续函数有效。要使用离散公式，我们只需对连续的高斯变换取样。这里使用离散变量表明我们正在处理取样后的变换。

是，$h(x)$有一个正的中心项及在其两侧的负项。显示在图 4.37(d)中的小模板“捕捉”了这一特点。这两个模板在第 3 章中用做锐化滤波，即我们目前所知道的高通滤波器。

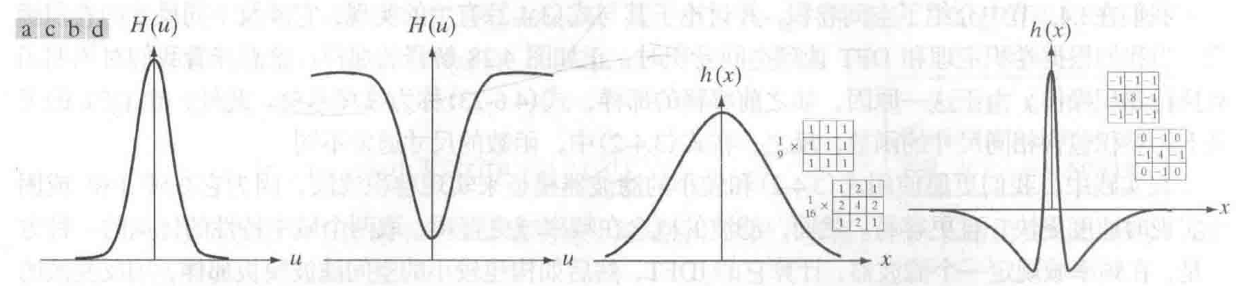


图 4.37 (a)频率域中的一维高斯低通滤波器；(b)对应于图(a)的空间低通滤波器；(c)频率域中的高斯高通滤波器；(d)对应于图(c)的空间高通滤波器。所示小二维模板是我们在第 3 章中用过的空间滤波器

虽然我们已经做了许多努力，但我们确信，如果没有刚刚建立的基础，要真正理解频率域滤波是不可能的。在实践中，频率域可被视为一个“实验室”，我们在这里取频率内容和图像外观之间相应的优点。如本章后面多次展示的那样，有些在空间域异常困难的或不可能直接用公式表达的任务，在频率域却变得很容易。一旦我们在频率域通过实验选择了一个特定的滤波器，那么该方法的实际实现通常是在空间域完成的。一种方法是指定试图捕获空间域中全部滤波器函数的“本质”的小空间模板，如我们在图4.37 中解释的那样。更正式的方法是，以数学或统计规则为基础，近似设计一个二维数字滤波器。4.11.4 节中将再次探讨这一主题。

例 4.15 由小空间模板得到频率域滤波器。

在这个例子中，我们从一个空间模板开始，说明如何产生相应的频率域滤波器。然后，我们比较用频率域滤波和空间技术得到的滤波结果。当希望比较给定空间模板相对于一个或多个频率域全滤波器的性能时，或深入理解模板的性能时，这类分析是很有用的。为简单起见，我们使用来自图 3.41(e)中的 3×3 Sobel 垂直边缘检测器。图 4.38(a)显示了一幅 600 × 600 像素的图像 $f(x, y)$，我们要对它进行滤波，图4.38(b)显示了该图像的谱。

图4.39(a)显示了这个 Sobel 模板 $h(x, y)$(透视图在下面解释)。因为输入图像的大小是 600×600 像素，而滤波器的大小为 3×3，为避免缠绕错误，我们根据式(4.6-29)和式(4.6-30)将 f 和 h 填充为 602×602 像素大小，Sobel 模板是奇对称的，它被嵌入到一个偶数尺寸的 0 阵列中(见例 4.10)。为保持这种对称性，我们把 $h(x, y)$的中心放在 602×602 这个填充后的阵列的中心处。对滤波器生成来说，这是一个很重要的方面。如果在形成 $h_p(x, y)$的过程中，填充后的阵列保持奇对称，则由表 4.1 中的性质 9 可知 $H(u, v)$将完全是虚函数。如本例末尾显示的那样，这将得到用 $h(x, y)$对图像进行空间滤波相同的结果。如果不保持这种对称性，结果将不再相同。

产生 $H(u, v)$的过程是：(1)用$(-1)^{x+y}$乘以 $h_p(x, y)$使频率域滤波器“中心化”；(2)计算(1)中结果的正 DFT；(3)考虑寄生的实部(我们知道 $H(u, v)$必须是纯虚函数)，将得到的 DFT 的实部置 0；(4)用$(-1)^{u+v}$乘以该结果。最后一步反过来用$(-1)^{u+v}$乘以 $H(u, v)$，它隐含着 $h(x, y)$被移到 $h_p(x, y)$的中心。图4.39(a)显示了 $H(u, v)$的透视图，而图4.39(b)以图像的方式显示了 $H(u, v)$。如期望的那样，该函数是奇函数，因而其关于中心反对称。在 4.7.3 节略述的步骤中，函数 $H(u, v)$被用做任何其他的频率域滤波器。

图4.39(c)是在 4.7.3 节略述步骤中使用刚才得到的滤波器对图4.38(a)中的图像滤波的结果。正如我们对推导出的滤波器所期望那样，边缘得到了增强，所有恒值灰度区域被减小为 0(为了显示，灰度色调进行了标定)。图4.39(d)显示了在 3.6.4 节给出的步骤中使用 $h(x, y)$直接在空间域对同一图像滤波的结果，结果相同。

288 ~ 291

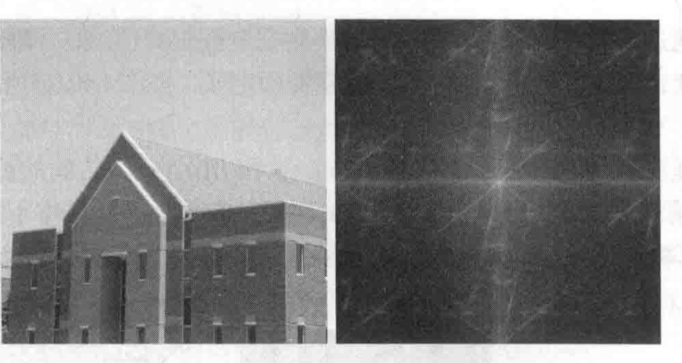


图 4.38　(a) 一幅建筑物的图像；(b) 该图像的谱

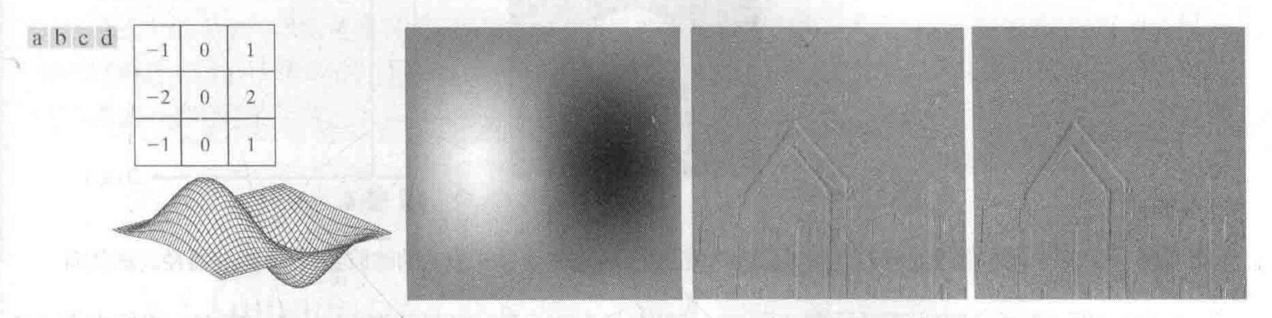


图 4.39　(a) 一个空间模板和它的频率域中对应的滤波器的透视图；(b) 以图像形式显示的滤波器；(c) 在频率域中使用图 (b) 中的滤波器对图 4.38 (a) 滤波后的结果；(d) 使用 (a) 中的空间滤波器对同一图像滤波的结果。结果相同

4.8　使用频率域滤波器平滑图像

本章的剩余部分将介绍频率域中的各种滤波技术，首先介绍低通滤波器。一幅图像中的边缘和其他急剧灰度变化（如噪声）主要影响其傅里叶变换的高频内容。因此，在频率域平滑（模糊）可通过衰减高频来达到，也就是使用低通滤波。本节介绍三种类型的低通滤波器：理想滤波器、布特沃斯滤波器和高斯滤波器。这三种滤波器涵盖了从非常急剧（理想）的滤波到非常平滑（高斯）的滤波范围。布特沃斯滤波器有一个参数，它称为滤波器的“阶数”。当阶数值较高时，布特沃斯滤波器接近于理想滤波器。对于较低的阶数值，布特沃斯滤波器更像高斯滤波器。这样，布特沃斯滤波器就可视为两种“极端”滤波器的过渡。本节中所有的滤波器都遵循 4.7.3 节概述的步骤，因此所有的滤波函数 $H(u, v)$ 可理解为大小为 $P \times Q$ 的离散函数，即离散频率变量的范围是 $u = 0, 1, 2, \cdots, P-1$ 和 $v = 0, 1, 2, \cdots, Q-1$。

4.8.1　理想低通滤波器

在以原点为圆心、以 D_0 为半径的圆内，无衰减地通过所有频率，而在该圆外“阻断”所有频率的二维低通滤波器，称为理想低通滤波器（ILPF）；它由下面的函数确定：

$$H(u,v) = \begin{cases} 1, & D(u,v) \le D_0 \\ 0, & D(u,v) > D_0 \end{cases} \tag{4.8-1}$$

式中，D_0 是一个正常数，$D(u, v)$ 是频率域中点 (u, v) 与频率矩形中心的距离，即

$$D(u,v) = \left[(u - P/2)^2 + (v - Q/2)^2 \right]^{1/2} \tag{4.8-2}$$

式中，与以前一样，P 和 Q 是用式 (4.6-31) 和式 (4.6-32) 填充后的尺寸。图 4.40 (a) 显示了 $H(u, v)$ 的透视图，图 4.40 (b) 显示了以图像方式呈现的滤波器。如 4.3.3 节提到的那样，理想一词表明在半径为 D_0 的圆

290 ~ 291

内，所有频率无衰减地通过，而在此圆之外的所有频率则完全被衰减(滤除)。理想低通滤波器是关于原点径向对称的，这意味着该滤波器完全由一个径向横截面定义，如图4.40(c)所示。将该横截面旋转360°则得到二维滤波器。

对于一个理想低通滤波器(ILPF)横截面，在 $H(u, v) = 1$ 和 $H(u, v) = 0$ 之间的过渡点称为截止频率。例如，在图4.40 所示的情况下，截止频率为 D_0。ILPF 的这种陡峭的截止频率不能使用电子元件来实现，尽管它们可以在计算机上仿真。对数字图像使用这些"非物理"滤波器的效果将在本节后面讨论。

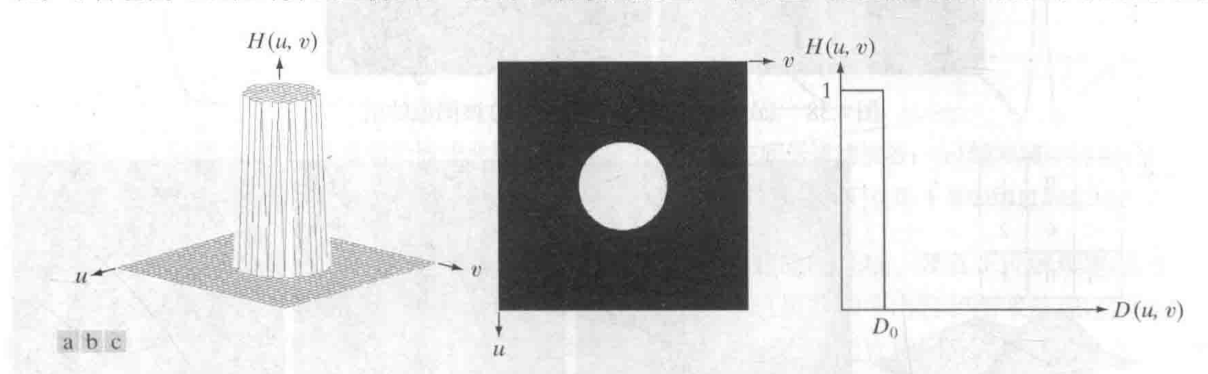


图 4.40　(a)一个理想低通滤波器变换函数的透视图；(b)以图像形式显示的滤波器；(c)滤波器径向横截面

本章介绍的低通滤波器可用具有相同截止频率的函数研究其特性而加以比较。建立一组标准截止频率轨迹的一种方法是，计算包含规定的总图像功率值 P_T 的圆。该值是通过求每个点(u, v)处填充后图像的功率谱成分之和得到的，其中 $u = 0, 1, \cdots, P-1$，$v = 0, 1, \cdots, Q-1$，即

$$P_T = \sum_{u=0}^{P-1}\sum_{v=0}^{Q-1} P(u,v) \tag{4.8-3}$$

其中，$P(u, v)$已在式(4.6-18)中给出。如果 DFT 已被中心化，那么原点位于频率矩形中心处、半径为 D_0 的圆将包含 α%的功率，其中

$$\alpha = 100\left[\sum_u\sum_v P(u,v)/P_T\right] \tag{4.8-4}$$

总和取处于圆之内或边界线上的(u, v)值。

图4.41(a)和(b)显示了一幅测试图像及其傅里叶谱。谱上重叠的圆的半径分别为 10, 30, 60, 160 和 460 像素。这些圆包围的图像功率(α%)分别为 87.0%, 93.1%, 95.7%, 97.8%和 99.2%。谱迅速衰落，总功率的 87%包含在半径为 10 的小圆内。



图 4.41　(a)大小为 688×688 像素的测试图像；(b)测试图像的傅里叶谱。由于填充，该谱的大小为图像大小的两倍，但是，它是以一半大小来显示的，因此，在页面上是匹配的。在谱上重叠的圆的半径相对于全尺寸的谱图像来说分别为 10, 30, 60, 160 和 460。这些半径分别包含了填充后图像功率的 87.0%, 93.1%, 95.7%, 97.8%和 99.2%

292 ~ 293

例 4.16　使用一个 ILPF 平滑图像。

图4.42 显示了应用图4.41(b)中所示半径的截止频率的 ILPF 的结果。图4.42(b)对所有的实际目的没有意义，除非模糊的目的是消除图像中的所有细节，但表示最大物体的“斑点”除外。此图像中的严重模糊表明，图像中多数尖锐的细节信息包含在被滤波器滤除的 13%的功率内。随着滤波器半径的增大，滤除的功率越来越少，导致的模糊也越来越弱。注意，图 4.42(c)到图 4.42(e)中的图像都有“振铃”特征，随着被滤除的高频内容的数量的减少，图像的纹理变得越来越好。甚至在只滤除了总功率的 2%的图像［见图 4.42(e)］中，振铃现象也很明显。正如稍后我们将要看到的那样，这种振铃现象是理想滤波器的一种特性。最后，$\alpha = 99.2\%$时的结果在噪声区域内有非常轻微的模糊，但大部分图像与原图像非常接近。这表明在这种特殊情况下，图像中很少有边缘信息包含在 0.8%以上的谱功率中。

从这个例子中可以清楚地看到，理想低通滤波器并不是非常实用。然而，作为滤波概念发展的一部分，研究这种滤波器的特性非常有用。同时，正如以下讨论所示，通过试图解释 ILPF 在空间域的振铃特性，可以得到一些有趣的见解。

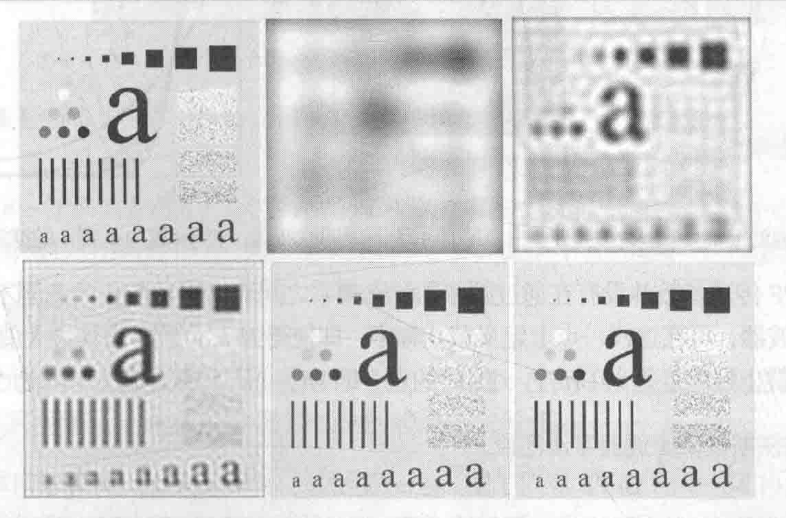


图 4.42　(a)原图像；(b)~(f)使用 ILPF 滤波后的结果，截止频率分别设置在半径值 10, 30, 60, 160 和 460 处，如图 4.41(b)所示。这些滤波器移除的功率分别为总功率的 13%, 6.9%, 4.3%, 2.2%和 0.8%

ILPF 的模糊和振铃特性可用卷积定理来解释。图 4.43(a)显示了半径为 10 的 ILPF 的空间表示 $h(x, y)$，图 4.43(b)显示了一条通过图像中心的水平线的灰度剖面线。由于 ILPF 在频率域的剖面图类似于盒状滤波器，因此可以预料相应空间滤波器具有 sinc 函数形状。空间域滤波可通过 $h(x, y)$与图像卷积来实现。将图像中的每个像素想象为一个离散冲激，它的强度与所在位置的灰度成正比。一个 sinc 函数与一个冲激卷积就是在冲激处复制这个 sinc 函数。sinc 函数的中心波瓣是引起模糊的主因，而外侧较小的波瓣是造成振铃的主要原因。sinc 函数与图像中的每个像素卷积为解释 ILPF 的特性提供了一个很好的模型。因为 sinc 函数的“展开度”与 $H(u, v)$的半径成反比，D_0越大，这个空间 sinc 函数就越趋近于一个与图像卷积时根本不会导致模糊的冲激。我们现在应已习惯了这种类型的矛盾特性。

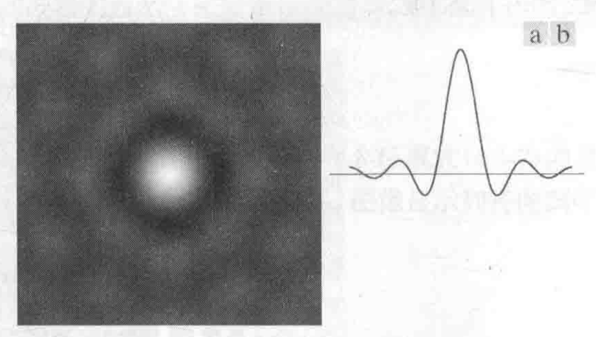


图 4.43　(a)半径为 5、大小为 1000×1000 的 ILPF 的空间域表示；(b)一条通过图像中心的水平线的灰度剖面图

293~295

在接下来的两节中，我们将说明很小或没有振铃而达到模糊是可能的，这是低通滤波的重要目标。

> 布特沃斯低通滤波器的传递函数通常写为此处表达式的平方根。由于此处我们关心的是这种滤波器的基本形式，因此为计算方便去掉了平方根。

4.8.2 布特沃斯低通滤波器

截止频率位于距原点 D_0 处的 n 阶布特沃斯低通滤波器(BLPF)的传递函数定义为

$$H(u,v)=\frac{1}{1+\left[D(u,v)/D_0\right]^{2n}} \tag{4.8-5}$$

式中，$D(u,v)$由式(4.8-2)给出。图 4.44 显示了该 BLPF 函数的透视图、图像显示和径向剖面图。

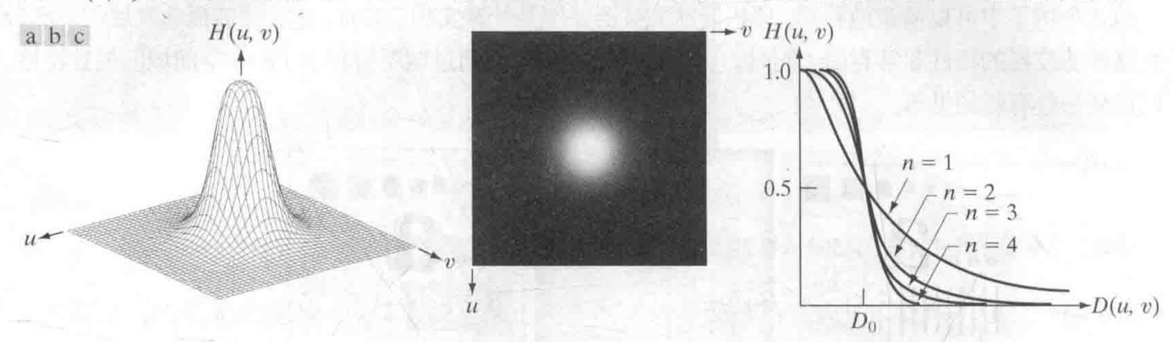


图 4.44 (a)布特沃斯低通滤波器传递函数的透视图；(b)显示为图像的滤波器；(c)阶数为 1 到 4 的滤波器的径向剖面

与 ILPF 不同，BLPF 传递函数并没有在通过频率和滤除频率之间给出明显截止的急剧不连续性。对于具有平滑传递函数的滤波器，可在这样一点上定义截止频率，即使得$H(u,v)$下降为其最大值的某个百分比的点。在式(4.8-5)中，截止频率点是当$D(u,v)=D_0$时的点［即$H(u,v)$从其最大值 1 下降为 50%］。

例 4.17 使用布特沃斯低通滤波器平滑图像。

图4.45 显示了将式(4.8-5)中的 BLPF 应用于图4.45(a)的结果，其中 $n=2$，D_0等于图4.41(b)所示的 5 个半径。不像图 4.42 所示的 ILPF 的结果，我们注意到，此处模糊的平滑过渡是截止频率增大的函数。此外，使用这种特殊 BLPF 处理过的任何图像中都没有可见的振铃现象，这要归因于这种滤波器在低频和高频之间的平滑过渡。

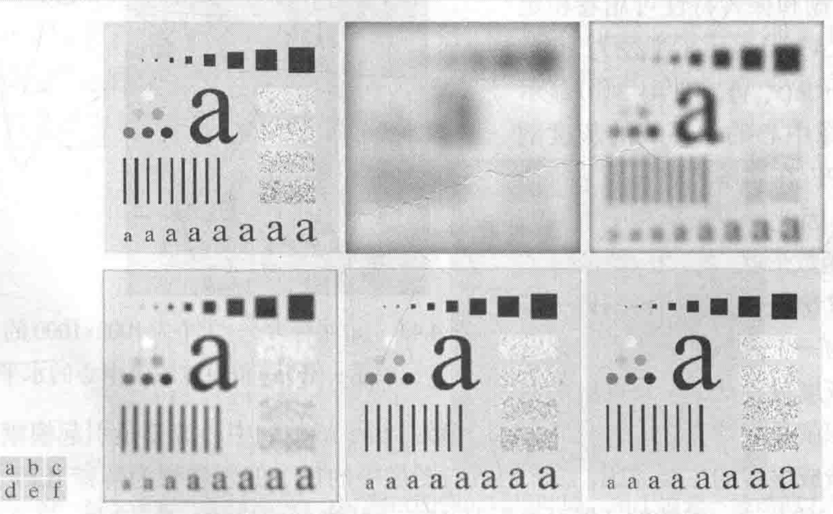


图 4.45 (a)原图像；(b)~(f)使用 2 阶 BLPF 的结果，BLPF 的截止频率位于图 4.41 所示的半径处。与图 4.42 比较

295 ~ 297

空间域的一阶布特沃斯滤波器没有振铃现象。在二阶滤波器中，振铃现象通常很难察觉，但更高阶数的滤波器中振铃现象会很明显。图4.46 比较了不同阶数的 BLPF 的空间表示(所有情况下均使用截止频率 5)，还显示了通过每个滤波器中心的水平扫描线的灰度剖面图。这些滤波器是使用生成图4.43 的相同过程来获得并显示的。为便于比较，使用伽马变换［见式(3.2-3)］对图4.46 中的图像进行了额外增强。一阶 BLPF［见图4.46(a)］没有振铃和负值。二阶 BLPF 的确显示了轻微的振铃和较小的负值，但远没有在 ILPF 中那样明显。正如其余图像所示，在 BLPF 中，高阶滤波器中的振铃现象变得很明显。20 阶布特沃斯滤波器呈现出了与 ILPF 类似的特性(极限情况下，两个滤波器相同)。因此，二阶 BLPF 是在有效的低通滤波和可接受的振铃特性之间的较好折中。

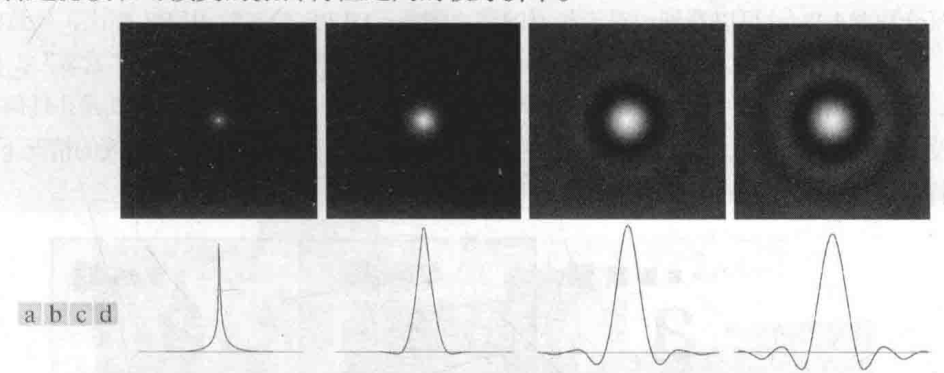


图 4.46　(a)~(d)阶数分别为 1, 2, 5 和 20 的 BLPF 的空间表示及相应的通过滤波器中心的灰度剖面图(所有情况下大小均为 1000×1000，且截止频率为 5)。观察作为滤波器阶数的函数的振铃是如何增加的

4.8.3　高斯低通滤波器

一维高斯低通滤波器(GLPF)在 4.7.4 节中已介绍过，当时介绍它的目的是用来帮助寻找空间域与频率域之间的重要联系。这些滤波器的二维形式由下式给出：

$$H(u,v)=\mathrm{e}^{-D^2(u,v)/2\sigma^2} \tag{4.8-6}$$

其中，如式(4.8-2)中那样，$D(u, v)$是距频率矩形中心的距离。为与本节讨论的滤波器保持一致，这里我们并未像在4.7.4节那样使用一个最高值为 1 的常量乘数。与以前一样，σ 是关于中心的扩展度的度量。通过令 $\sigma=D_0$，我们可以在本节中使用其他滤波器的表示法来表示该滤波器：

$$H(u,v)=\mathrm{e}^{-D^2(u,v)/2D_0^2} \tag{4.8-7}$$

式中，D_0是截止频率。当 $D(u, v)=D_0$时，GLPF 下降到其最大值的 0.607 处。

如表 4.3 所示，GLPF 的傅里叶反变换也是高斯的。这意味着通过计算式(4.8-6)或式(4.8-7)的 IDFT 得到的空间高斯滤波器将没有振铃。图4.47 显示了一个 GLPF 函数的透视图、图像显示和径向剖面图，表 4.4 总结了本节讨论的低通滤波器。

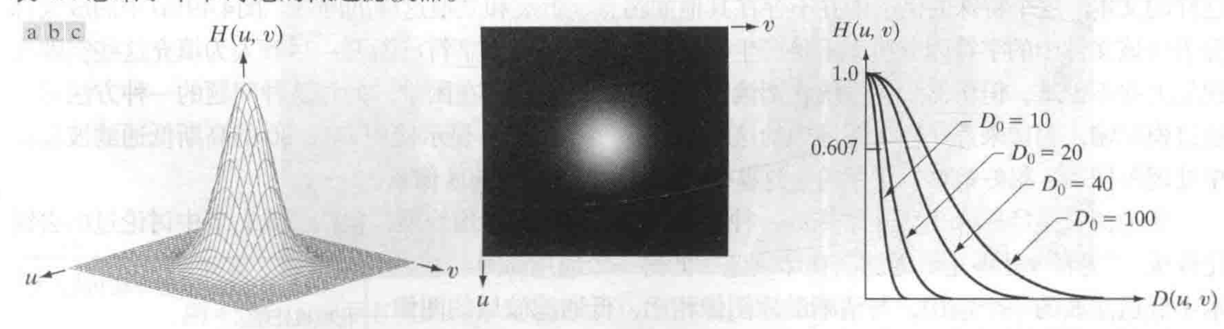


图 4.47　(a)一个 GLPF 传递函数的透视图；(b)显示为图像的滤波器；(c)不同 D_0 值的滤波器径向剖面图

296~299

表 4.4 低通滤波器。D_0 是截止频率，n 是布特沃斯滤波器的阶数

理 想	布特沃斯	高 斯
$H(u,v)=\begin{cases}1, & D(u,v)\leqslant D_0\\ 0, & D(u,v)>D_0\end{cases}$	$H(u,v)=\dfrac{1}{1+[D(u,v)/D_0]^{2n}}$	$H(u,v)=\mathrm{e}^{-D^2(u,v)/2D_0^2}$

例 4.18 使用高斯低通滤波器平滑图像。

图 4.48 显示了将式(4.8-7)给出的 GLPF 应用到图 4.48(a)后的结果，其中 D_0 等于图 4.41(b)中的 5 个半径。正如二阶 BLPF(见图 4.45)的情况那样，我们注意到模糊的平滑过渡是截止频率增大的函数。比较图 4.45(c)和图4.48(c)可以看到，对于相同的截止频率，GLPF 与二阶 BLPF 相比，导致的平滑效果要稍差一些。这正如我们所料，因为 GLPF 的剖面图不像二阶 BLPF 的剖面图那样“紧凑”。但结果还是可以比较的，且我们可以确信在 GLPF 中没有振铃。这是实际中的一个重要特性，尤其是在任何类型的人工缺陷不可接受的情况(如医学成像)下。在需要严格控制低频和高频之间截止频率的过渡的情况下，BLPF 是更合适的选择。额外控制滤波器剖面线的代价是可能会产生振铃。

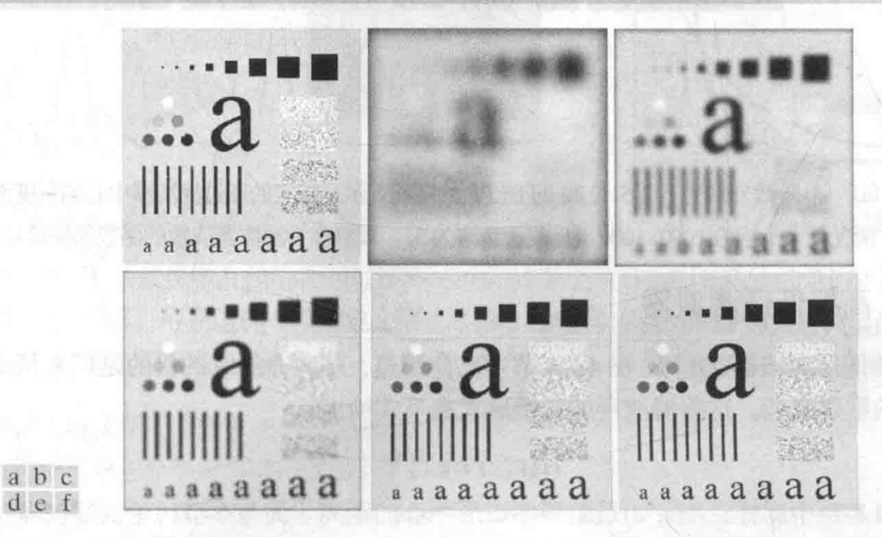


图 4.48 (a)原图像；(b)~(f)使用截止频率位于图 4.41 所示半径处的 GLPF 对图像滤波后的结果。请比较图 4.42 和图 4.45

4.8.4 低通滤波的其他例子

在下面的讨论中，我们将说明频率域中低通滤波的几个实际应用。第一个例子来自机器感知领域，即字符识别的应用；第二个例子来自印刷和出版业；第三个例子关系到卫星图像和航空图像的处理。类似的结果可用 3.5 节讨论的低通空间滤波技术得到。

图4.49 显示了一幅低分辨率文本图像的例子。例如，我们会在传真、复印材料和历史记录中遇到这样的文本。这个特殊的例子中并不存在其他如污点、折痕和破痕这样的问题。图4.49(a)中的放大部分表明该文件中的字符因分辨率不足产生了形状失真，且许多字符已断开。尽管人为填充这些空隙在视觉上并不困难，但机器识别系统在阅读这些断开字符时却存在困难。处理这种问题的一种方法是，通过模糊输入图像来连接输入图像中的这些小空缝。图4.49(b)显示使用 $D_0 = 80$ 的高斯低通滤波器简单处理图像后，很好地修复了字符。这些图像的大小为 444 × 508 像素。

低通滤波是印刷和出版业常用的一种技术，在这里用来做预处理，包括 3.6.3 节中讨论过的去锐化模板。“美容”处理是低通滤波在印刷业中的另一个应用。图4.50 显示了低通滤波的一个应用，与清晰的原图像相比，低通滤波后的图像看上去更平滑、更柔和。对于人脸，主要目的是减少皮肤细纹的锐化

4.9.5 节中将讨论频率域中的去锐化模板处理。

299 ~ 301

程度和小斑点。从图4.50(b)和图4.50(c)中的放大部分可明显看出，眼睛周围的细纹已明显减少。事实上，平滑后的图像看上去更为柔和、更美观。

a b

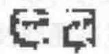


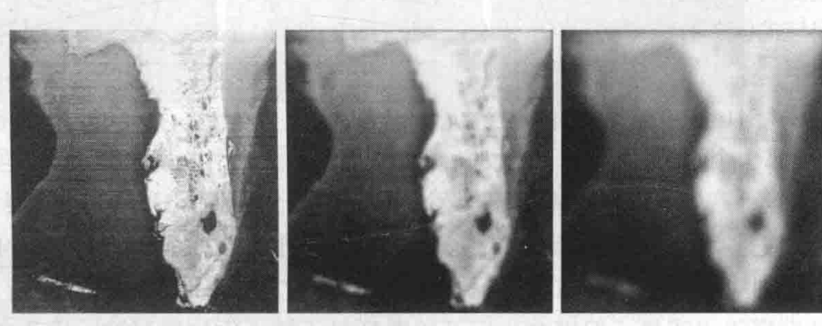
图 4.49 (a)低分辨率文本例子(注意放大图中的断开字符)；(b)使用 GLPF 对图像滤波后的结果(断开的字符段已被连在一起)



a b c

图 4.50 (a)原图像(784×732 像素)；(b)使用 $D_0=100$ 的 GLPF 对图像滤波后的结果；(c)使用 $D_0=80$ 的 GLPF 对图像滤波后的结果。注意图(b)和图(c)中放大部分的皮肤皱纹已减少了

图4.51显示了相同图像的两种低通滤波应用，但是具有不同目的。图4.51(a)是一幅显示了部分墨西哥湾(暗色)和佛罗里达(亮色)的、大小为 808×754 的甚高分辨率辐射计(VHRR)图像，该图像由 NOAA 卫星拍摄(注意水平传感器扫描线)。水体间的边界是由环流引起的。这幅图像是直观的遥感图像，传感器具有沿场景扫描方向产生突出扫描线的倾向(见例 4.24 中关于物理原因的说明)。低通滤波是降低这些扫描线的影响的简略方法，如图 4.51(b)所示(4.10 节和 5.4.1 节中将考虑更为有效的方法)。该图像是使用 $D_0=50$ 的高斯低通滤波器得到的。降低扫描线的影响可简化如洋流间边界特征的检测。



a b c

图 4.51 (a)显示有突出扫描线的图像；(b)使用 $D_0=50$ 的 GLPF 对图像滤波后的结果；(c)使用 $D_0=20$ 的 GLPF 对图像滤波后的结果(原图像由 NOAA 提供)

301 ~ 303

图4.51(c)显示了使用 $D_0 = 20$ 的高斯低通滤波器对图像滤波更有意义的结果。这里，对图像滤波的目的是尽可能模糊更多的细节，而保留大的可识别特征。例如，这种类型的滤波可以作为图像分析系统的预处理的一部分，该系统在图像库中搜索特征。具有这种特征的一个例子可能是给定大小的湖泊，例如佛罗里达东南部的奥基乔比湖——图 4.51(c)中的一个近似环形的暗区。低通滤波通过消除比感兴趣的平均特征小的其他特征来简化分析。

4.9 使用频率域滤波器锐化图像

前几节中说明了通过衰减图像的傅里叶变换的高频成分可以平滑图像。因为边缘和其他灰度的急剧变化与高频成分有关，所以图像的锐化可在频率域通过高通滤波来实现，高通滤波会衰减傅里叶变换中的低频成分而不会扰乱高频信息。如 4.8 节那样，我们仅考虑径向对称的零相移滤波器。本节中的所有滤波都以 4.7.3 节概述的过程为基础，因此所有的滤波器函数 $H(u, v)$都可理解为大小为 $P\times Q$ 的离散函数，即离散频率变量的范围是 $u = 0, 1, 2, \cdots, P-1$ 和 $v = 0, 1, 2, \cdots, Q-1$。

一个高通滤波器是从给定低通滤波器用下式得到的：

$$H_{\mathrm{HP}}(u,v) = 1 - H_{\mathrm{LP}}(u,v) \tag{4.9-1}$$

式中，$H_{\mathrm{LP}}(u, v)$是低通滤波器的传递函数。也就是说，被低通滤波器衰减的频率能通过高通滤波器，反之亦然。

本节考虑理想、布特沃斯和高斯高通滤波器。如前几节所示，我们在频率域和空间域分别说明这些滤波器的特性。图4.52 显示了这些滤波器的典型三维图、图像表示和剖面图。像前面一样，我们看到，布特沃斯滤波器表现为理想滤波器的锐利性和高斯滤波器的宽阔平滑性间的一种过渡。紧随该节后面的讨论，图4.53 说明了这些滤波器在空间域的形状。空间滤波器是使用用于生成图 4.43 和图 4.46 的步骤得到和显示的。

4.9.1 理想高通滤波器

二维理想高通滤波器(IHPF)定义为

$$H(u,v) = \begin{cases} 0, & D(u,v) \leqslant D_0 \\ 1, & D(u,v) > D_0 \end{cases} \tag{4.9-2}$$

式中，D_0是截止频率，$D(u, v)$由式(4.8-2)给出。这个表达式直接来自式(4.8-1)和式(4.9-1)。如所预料的那样，IHPF 与 ILPF 是相对的，在这种情况下，IHPF 把以半径为 D_0的圆内的所有频率置零，而毫无衰减地通过圆外的所有频率。如同 ILPF 那样，IHPF 也是物理上无法实现的。但与前面一样，我们这里出于完整性对它进行讨论，因为 IHPF 可以用来解释空间域中的振铃等现象，但讨论很简要。

因为它们与式(4.9-1)的方法有关，我们可以预料 IHPF 和 ILPF 一样具有相同的振铃性质，这清楚地展示在图 4.54 中，这些图像是使用 D_0分别为 30, 60 和 160 像素的各种 IHPF 对图 4.41(a)中的原始图像滤波后的结果。图 4.54(a)中的振铃现象相当严重，以至于产生了失真，物体的边界也被加粗(看一下那个大字母“a”)。顶部 3 个圆的显示也不是很好，因为它们不像图像中的其他边缘那样强(这 3 个目标的灰度很接近于背景灰度，产生了较小的幅度不连续)。观察图 4.53(a)中 IHPF 的空间表示的“点”的大小，记住，空间域滤波是空间滤波器与图像的卷积，这一概念可以帮助解释为什么较小的物体和线条几乎显示为纯白色。特别应注意图 4.54(a)中最上一行的 3 个小方形和垂直细条带。当 $D_0 = 60$ 时，情况有所改善。边缘失真仍很明显，但现在我们开始看到对于较小物体的过滤。由于现在已很熟悉频率域和空间域的相反关系，我们知道这个滤波器的点的大小要比 $D_0 = 30$ 的滤波器更小。$D_0 = 160$ 的结果

302 ~ 306

更像一幅高通滤波后的图像。图中，边缘更清晰，失真更小，而且较小的物体已被正确滤除。当然，所有图像的恒定背景在高通滤波后的图像中都是 0，因为高通滤波类似于空间域中的差分。

图 4.52　顶部行：典型理想高通滤波器的透视图、图像表示和剖面图；中间行：典型布特沃斯高通滤波器的透视图、图像表示和剖面图和；底部行：典型高斯高通滤波器的透视图、图像表示和剖面图

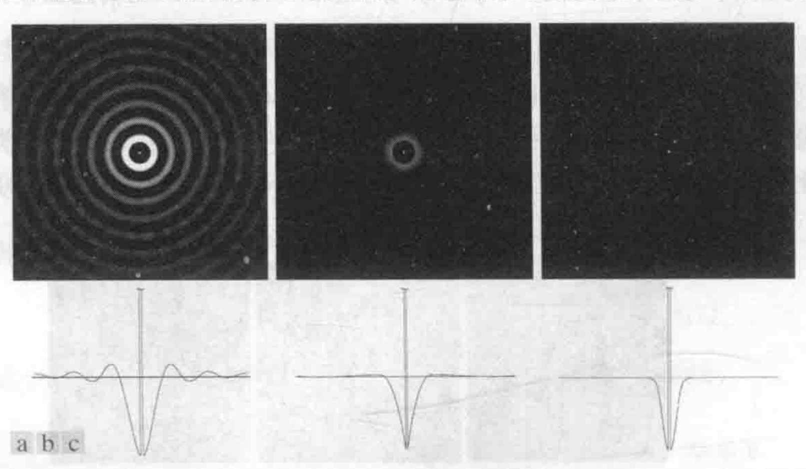


图 4.53　(a)典型的频率域理想高通滤波器的空间表示及通过滤波器中心的对应灰度剖面图；(b)典型的频率域布特沃斯高通滤波器的空间表示及通过滤波器中心的对应灰度剖面图；(c)典型的高斯频率域高通滤波器的空间表示及通过滤波器中心的对应灰度剖面图

304~306

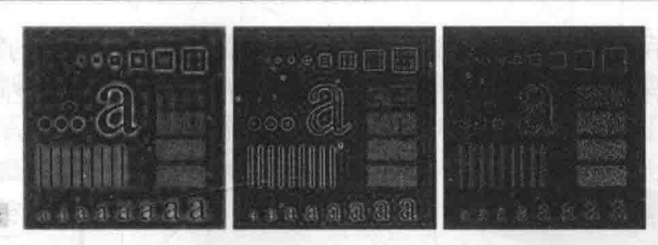


图 4.54 使用 $D_0 = 30, 60$ 和 160 的 IHPF 对图 4.41(a) 中的图像滤波后的结果

4.9.2 布特沃斯高通滤波器

截止频率为 D_0 的 n 阶布特沃斯高通滤波器(BHPF)定义为

$$H(u,v) = \frac{1}{1+\left[D_0 / D(u,v)\right]^{2n}} \tag{4.9-3}$$

其中 $D(u, v)$由式(4.8-2)给出。该表达式直接来自式(4.8-5)和式(4.9-1)。图 4.52 的中间一行显示了该 BHPF 函数的图像和剖面图。

与低通滤波器一样，我们可以预料布特沃斯高通滤波器比 IHPF 更平滑。图 4.55 显示了 D_0 值与图 4.54 中设置相同的 2 阶 BHPF 的特性。边缘失真比图 4.54 小得多，甚至对于最小的截止频率值也一样。因为 IHPF 和 BHPF 中心区域的点的大小类似［见图4.53(a)和图4.53(b)］，这两个滤波器过滤小目标的性能可以比拟。截止频率越大，使用 BHPF 得到的结果就越平滑。

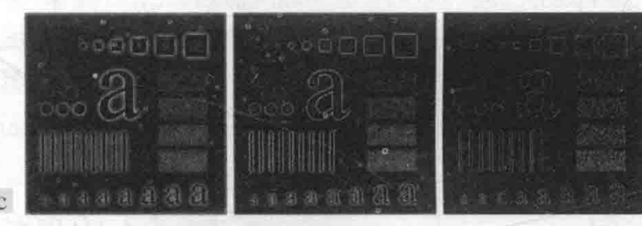


图 4.55 使用［对应于图 4.41(b) 中的］$D_0 = 30, 60$ 和 160 的 2 阶 BHPF 对图 4.41(a) 中的图像高通滤波后的结果。这些结果比用 IHPF 得到的结果平滑得多

4.9.3 高斯高通滤波器

截止频率处在距频率矩形中心距离为 D_0 的高斯高通滤波器(GHPF)的传递函数，由下式给出：

$$H(u,v) = 1 - \mathrm{e}^{-D^2(u,v)/2D_0^2} \tag{4.9-4}$$

其中，$D(u, v)$由式(4.8-2)给出。该表达式直接来自式(4.8-7)和式(4.9-1)。图4.52的第三行显示了 GHPF 函数的透视图、图像和剖面。沿用 BHPF 相同的形式，图4.56中显示了使用 GHPF 的类似结果。正如所料，得到的结果比前两个滤波器的结果更平滑。即使是对微小物体和细线条使用高斯滤波器滤波，结果也是较清晰的。表 4.5 总结了本节讨论的高通滤波器。

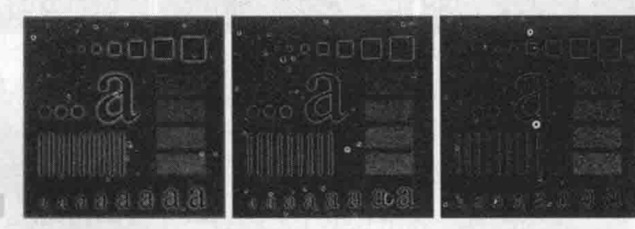


图 4.56 使用 $D_0 = 30, 60$ 和 160 的一个 GHPF 对图 4.41(a) 中的图像进行高通滤波后的结果，D_0 对应于图 4.41(b) 中的各个圆圈。请与图 4.54 和图 4.55 进行比较

305 ~ 308

表 4.5　高通滤波器。D_0 是截止频率，n 是布特沃斯滤波器的阶数

理　想	布特沃斯	高　斯
$H(u,v)=\begin{cases}0, & D(u,v)\leqslant D_0\\ 1, & D(u,v)>D_0\end{cases}$	$H(u,v)=\dfrac{1}{1+[D_0/D(u,v)]^{2n}}$	$H(u,v)=1-\mathrm{e}^{-D^2(u,v)/2D_0^2}$

例 4.19　使用高通滤波法和阈值法增强图像。

图4.57(a)是一幅大小为 1026 × 962 的指纹图像，其污染(一种典型的问题)是很明显的。自动指纹识别的关键一步是增强指纹的脊线并减小污染。增强在人对指纹的解释方面也很有用。在这个例子中，我们使用高通滤波法来增强脊线并减小污染效应。脊线的增强使用脊线包含高频这样一个事实来完成，而高通滤波器不会改变高频。另一方面，滤波器减少了低频成分，而低频成分对应图像中变化缓慢的灰度，例如背景和污染。这样，增强就可通过减少除高频外的所有特征来达到，在这种情况下，高频是感兴趣的特征。

图4.57(b)是使用截止频率为 50 的 4 阶布特沃斯高通滤波器滤波后的结果。如期望的那样，高通滤波后的图像失去了它的灰度色调，因为直流项已被减小为 0。最终结果是，高通滤波后的图像中，典型的暗色调已成为主流，因此需要额外的处理来增强感兴趣的细节。一种简单的方法是对滤波后的图像进行阈值处理。图4.57(c)显示了将滤波后的图像中的所有负值设置为黑、将所有正值设置为白后的结果。注意，脊线已变清晰，且污染明显减小。事实上，图4.57(a)所示图像右上方部分中几乎已看不见的脊线，在图4.57(c)中已被很好地增强。

> 值 D_0 大约为填充后图像的短维的 2.5%。D_0 接近于原点较为理想，因为这样会衰减低频，但低频不会完全消除。D_0 在短维的 2%至 5%是一个不错的起点。

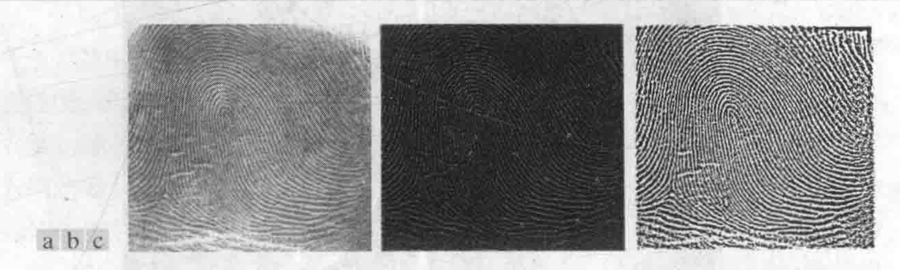


图 4.57　(a)指纹图像；(b)对图(a)高通滤波后的结果；(c)对图(b)进行阈值操作后的结果(原图像由美国国家标准和技术研究所提供)

4.9.4　频率域的拉普拉斯算子

在3.6.2节，我们已使用拉普拉斯算子对空间域图像进行了增强。本节重温拉普拉斯算子，并说明它可得到与频率域技术的等效结果。可以证明(见习题 4.26)，拉普拉斯算子能使用如下滤波器在频率域实现：

$$H(u,v)=-4\pi^2(u^2+v^2) \tag{4.9-5}$$

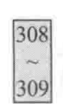
或者，关于频率矩形的中心，使用如下滤波器实现：

$$H(u,v)=-4\pi^2\left[(u-P/2)^2+(v-Q/2)^2\right]=-4\pi^2D^2(u,v) \tag{4.9-6}$$

式中，$D(u, v)$是式(4.8-2)给出的距离函数。然后，拉普拉斯图像由下式得到：

$$\nabla^2 f(x,y)=\Im^{-1}\left\{H(u,v)F(u,v)\right\} \tag{4.9-7}$$

式中，$F(u, v)$是$f(x, y)$的傅里叶变换。如 3.6.2 节描述的那样，增强可使用下式实现：



$$g(x,y)=f(x,y)+c\nabla^2 f(x,y) \tag{4.9-8}$$

这里，$c=-1$，因为$H(u,v)$是负的。在第3章中，$f(x,y)$和$\nabla^2 f(x,y)$具有可比的数值。然而，用式(4.9-7)计算$\nabla^2 f(x,y)$会引入已标定系数的DFT，这些系数的幅度与f的最大值相比可能要大几个量级。这样，f与其拉普拉斯间的差别一定会被带入可比的范围。处理该问题的简单方法是，在计算$f(x,y)$的DFT之前，将$f(x,y)$的值归一化到区间[0, 1]内，并用它的最大值除$\nabla^2 f(x,y)$，进而将它带到近似区间[-1, 1]内(回忆可知拉普拉斯算子有负值)。然后，可以应用式(4.9-8)。

在频率域，式(4.9-8)可写为

$$g(x,y)=\Im^{-1}\{F(u,v)-H(u,v)F(u,v)\}=\Im^{-1}\{[1-H(u,v)]F(u,v)\}=\Im^{-1}\{[1+4\pi^2 D^2(u,v)]F(u,v)\} \tag{4.9-9}$$

虽然这一结果是精致的，但它也存在刚刚提到的标定问题，并且归一化系数不容易计算。由于这一原因，式(4.9-8)优先选择在频率域中实现，使用式(4.9-7)计算$\nabla^2 f(x,y)$，并用前节提到的方法来标定。

例4.20 使用拉普拉斯算子在频率域锐化图像。

图4.58(a)是与图3.38(a)相同的图像，图4.58(b)显示了使用式(4.9-8)后的结果，其中拉普拉斯算子在频率域用式(4.9-7)计算。标定使用与公式有关的描述来完成。比较图4.58(b)和图3.38(e)，我们可以看到，频率域和空间域的结果在视觉上是相同的。很明显，在这两幅图中，对应于图3.37(b)的拉普拉斯模板在中心有一个系数-8(习题4.26)。

a b

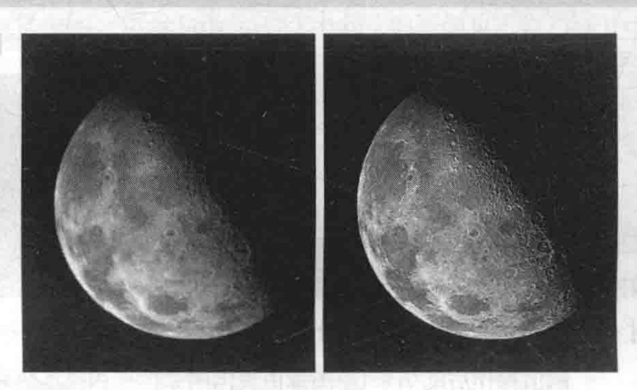


图4.58 (a)原始模糊图像；(b)在频率域中使用拉普拉斯算子增强后的图像。请与图3.38(e)比较

4.9.5 钝化模板、高提升滤波和高频强调滤波

本节讨论3.6.3节中介绍过的钝化模板、高提升滤波图像锐化技术。使用频率域方法，式(3.6-8)中定义的模板由下式给出：

$$g_{\text{mask}}(x,y)=f(x,y)-f_{\text{LP}}(x,y) \tag{4.9-10}$$

与

$$f_{\text{LP}}(x,y)=\Im^{-1}\left[H_{\text{LP}}(u,v)F(u,v)\right] \tag{4.9-11}$$

式中，$H_{\text{LP}}(u,v)$是一个低通滤波器，$F(u,v)$是$f(x,y)$的傅里叶变换。这里，$f_{\text{LP}}(x,y)$是平滑后的图像，该图像类似于式(3.6-8)中的$\overline{f}(x,y)$。然后，如式(3.6-9)那样，有

$$g(x,y)=f(x,y)+k*g_{\text{mask}}(x,y) \tag{4.9-12}$$

309 ~ 310

该表达式定义了$k=1$时的钝化模板和$k>1$时的高提升滤波器。利用前面的结果，我们完全可以使用涉及低通滤波器的频率域计算来表达式(4.9-12)：

$$g(x,y)=\Im^{-1}\left\{\left[1+k*\left[1-H_{\mathrm{LP}}(u,v)\right]\right]F(u,v)\right\} \tag{4.9-13}$$

用式(4.9-1)，我们可以根据高通滤波器来表达该结果：

$$g(x,y)=\Im^{-1}\left\{\left[1+k*H_{\mathrm{HP}}(u,v)\right]F(u,v)\right\} \tag{4.9-14}$$

方括号中的表达式称为高频强调滤波器。如前面提及的那样，高通滤波器将直流项置 0，因此会把滤波后的图像的平均灰度减小为 0。高频强调滤波器不存在这一问题，因为高通滤波器上加了 1。常数 k 给出了影响最终结果的高频的比例。高频强调滤波的通用公式为

$$g(x,y)=\Im^{-1}\left\{\left[k_1+k_2*H_{\mathrm{HP}}(u,v)\right]F(u,v)\right\} \tag{4.9-15}$$

其中，$k_1\geqslant0$ 控制距原点的偏移量［见图4.31(c)］，$k_2\geqslant0$ 控制高频的贡献。

例 4.21　使用高频强调滤波增强图像。

图4.59(a)显示了一幅大小为 416×596 的胸部 X 射线图像，其灰度级很窄。本例的目的是使用高频强调滤波增强该图像。由于 X 射线不能使用光学镜头聚焦，因此得到的图像通常都会稍微有些模糊。因为这幅特殊图像的灰度偏向灰度级的暗端，所以我们还利用这个机会给出一个如何使用空间域处理来补充频率域滤波的例子。

图4.59(b)显示了使用 $D_0=40$(大约为填充后图像大小的 5%)的一个高斯滤波器对图像进行高通滤波后的结果。如预料的那样，滤波的结果没有什么特征，但朦胧地显示了图像中的主边缘。图 4.95(c)显示了高频强调滤波的优点，其中我们使用了式(4.9-15)，且令 $k_1=0.5$，$k_2=0.75$。虽然图像依旧较暗，但其灰度级色调归因于低频成分而未丢失。

在医学图像处理中，像振铃这样的人为缺陷是不能接受的。因此，最好避免使用可能会在处理后的图像中引入人为缺陷的滤波器。因为空间和频率域高斯滤波器是傅里叶变换对，因此这些滤波器会产生不会引入人为缺陷的平滑结果。

如 3.3.1 节讨论的那样，在较窄范围的灰度级中，由灰度级表征的图像是直方图均衡处理的理想候选者。如图4.59(d)所示，这的确是进一步增强图像的合适方法。注意骨骼结构的清晰度和其他三幅图像中均不可见的细节。最终增强后的图像有点噪声，但在扩展图像的灰度级后，这是X光图像的典型现象。用高频强调和直方图均衡相结合得到的结果，优于用任何单独方法得到的结果。

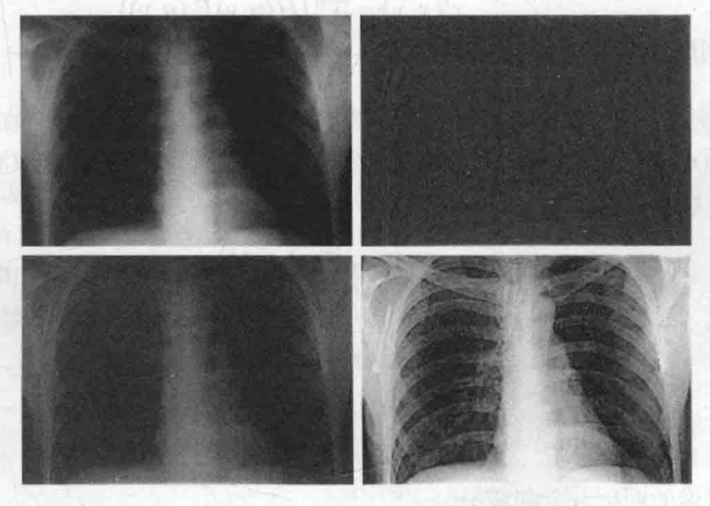


图 4.59　(a)一幅胸部 X 射线图像；(b)使用高斯高通滤波器滤波后的结果。(c)使用相同滤波器进行高频强调滤波后的结果；(d)对图(c)执行直方图均衡操作后的结果(原图像由密歇根大学医学院解剖科学部的 Thomas R. Gest 博士提供)

310 ~ 311

4.9.6 同态滤波

2.3.4 节中介绍的照射-反射模型可用于开发一种频率域处理过程，该过程通过同时压缩灰度范围和增强对比度来改善一幅图像的外观。由 2.3.4 节的讨论可知，一幅图像$f(x, y)$可以表示为其照射成分$i(x, y)$和反射成分 $r(x, y)$的乘积，即

$$f(x,y)=i(x,y)r(x,y) \tag{4.9-16}$$

上式不能直接用于对照射和反射的频率成分进行操作，因为乘积的傅里叶变换不是变换的乘积：

$$\Im[f(x,y)]\neq\Im[i(x,y)]\Im[r(x,y)] \tag{4.9-17}$$

然而，假定我们定义

> 如果灰度范围为$[0, L-1]$的图像$f(x, y)$出现任何 0 值，则图像中的每个元素必须加 1，以避免处理 ln(0) 的情形。在滤波处理结束时，再将 1 减去。

$$z(x,y)=\ln f(x,y)=\ln\ i(x,y)+\ln r(x,y) \tag{4.9-18}$$

则有

$$\Im\{z(x,y)\}=\Im\{\ln f(x,y)\}=\Im\{\ln\ i(x,y)\}+\Im\{\ln\ r(x,y)\} \tag{4.9-19}$$

或

$$Z(u,v)=F_i(u,v)+F_r(u,v) \tag{4.9-20}$$

式中，$F_i(u, v)$和 $F_r(u, v)$分别是 $\ln i(x, y)$和 $\ln r(x, y)$的傅里叶变换。

我们可以用滤波器 $H(u, v)$对 $Z(u, v)$滤波，故有

$$S(u,v)=H(u,v)Z(u,v)=H(u,v)F_i(u,v)+H(u,v)F_r(u,v) \tag{4.9-21}$$

在空间域中，滤波后的图像是

$$s(x,y)=\Im^{-1}\{S(u,v)\}=\Im^{-1}\{H(u,v)F_i(u,v)\}+\Im^{-1}\{H(u,v)F_r(u,v)\} \tag{4.9-22}$$

由定义

$$i'(x,y)=\Im^{-1}\{H(u,v)F_i(u,v)\} \tag{4.9-23}$$

和

$$r'(x,y)=\Im^{-1}\{H(u,v)F_r(u,v)\} \tag{4.9-24}$$

我们可以下列形式来表达式(4.9-23)：

$$s(x,y)=i'(x,y)+r'(x,y) \tag{4.9-25}$$

最后，因为 $z(x, y)$是通过取输入图像的自然对数形成的，我们可通过取滤波后的结果的指数这一反处理来形成输出图像：

$$g(x,y)=\mathrm{e}^{s(x,y)}=\mathrm{e}^{i'(x,y)}\mathrm{e}^{r'(x,y)}=i_0(x,y)r_0(x,y) \tag{4.9-26}$$

其中，

$$i_0(x,y)=\mathrm{e}^{i'(x,y)} \tag{4.9-27}$$

和

$$r_0(x,y)=\mathrm{e}^{r'(x,y)} \tag{4.9-28}$$

是输出(处理后)图像的照射和反射成分。

311 ~ 314

刚刚推导的滤波方法总结在图4.60 中。该方法是以称之为同态系统的一类系统的特殊情况为基础的。在这种特殊应用中，方法的关键在于照射成分和反射成分的分离，其实现形式如式(4.9-20)所示。然后，如式(4.9-21)指出的那样，同态滤波函数 $H(u, v)$可分别对这些成分进行操作。

$f(x,y)$ ⇨ ln ⇨ DFT ⇨ $H(u,v)$ ⇨ $(\text{DFT})^{-1}$ ⇨ exp ⇨ $g(x,y)$

图 4.60　同态滤波步骤小结

图像的照射成分通常由慢空间变化来表征，而反射成分往往引起突变，特别是在不同物体的连接处。这些特性导致图像取对数后的傅里叶变换的低频成分与照射相联系，而高频成分与反射相联系。虽然这些联系只是粗略的近似，但它们用在图像滤波中是有益的，如例 4.22 中说明的那样。

使用同态滤波器可更好地控制照射成分和反射成分。这种控制需要指定一个滤波器函数 $H(u,v)$，它可用不同的可控方法影响傅里叶变换的低频和高频成分。图 4.61 显示了这种滤波器的剖面图。如果γ_L和γ_H选定，而$\gamma_L<1$且$\gamma_H>1$，那么图4.61所示滤波器函数趋向于衰减低频（照射）的贡献，而增强高频（反射）的贡献。最终结果是同时进行动态范围的压缩和对比度的增强。

图4.61所示的函数形状可用高通滤波器的基本形式来近似。例如，采用形式稍微变化一下的高斯高通滤波器可得到函数

$$H(u,v)=(\gamma_H-\gamma_L)\left[1-\mathrm{e}^{-c\left[D^2(u,v)/D_0^2\right]}\right]+\gamma_L \tag{4.9-29}$$

其中，$D(u,v)$由式(4.8-2)定义，常数 c 控制函数边坡的锐利度，它在γ_L和γ_H之间过渡。这个滤波器类似于前节讨论的高频强调滤波器。

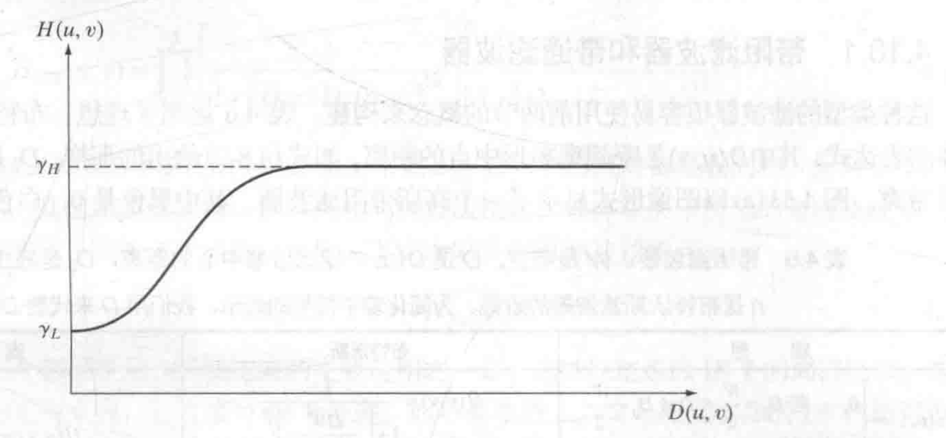


图 4.61　圆形对称同态滤波器函数的径向剖面图。垂直轴位于频率矩形的中心，$D(u,v)$是距中心的距离

例 4.22　使用同态滤波增强图像。

图 4.62(a)显示了一幅大小为 1162×746 像素的全身 PET (Positron Emission Tomography) 扫描图像。图像稍微有点儿模糊，且由于支配显示动态范围的高灰度"热点"使得许多低灰度特征很朦胧(这些"热点"是由脑部和肺部的肿瘤导致的)。图 4.62(b)是对图 4.62(a)使用同态滤波得到的，使用式(4.9-29)中的滤波器，$\gamma_L=0.25,\ \gamma_H=2,\ c=1,\ D_0=80$。该滤波器的剖面看上去正好像图 4.61，只是边坡稍微有点陡峭。

> 回忆可知滤波操作使用到了图像填充，因此该滤波器的大小为 $P\times Q$。

注意，如图4.62(b)所示，在处理后的图像中，"热点"、脑部和骨架更为清晰，且图像中可以见到更多的细节。对于显示的动态范围来说，通过降低支配照射成分(热点)的影响，允许低灰度变得更可见是可能的。类似地，因为高频被同态滤波增强，故图像的反射成分(边缘信息)被明显锐化。图 4.62(b)中增强后的图像较之原图像已有很大改进。

314 ~ 315

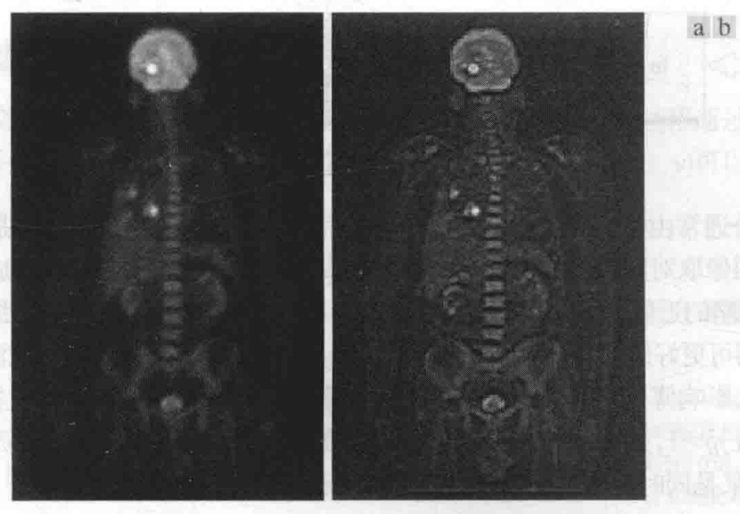


图 4.62 (a)全身 PET 扫描图像；(b)用同态滤波增强的图像(原图像由 CTI PET Systems 公司的 Michael E. Casey 博士提供)

4.10 选择性滤波

前两节讨论的滤波器在整个频率矩形上操作。在很多应用中，它们的目的是处理指定频段或频率矩形的小区域。第一类滤波器分别称为带阻滤波器或带通滤波器，第二类滤波器称为陷波滤波器。

4.10.1 带阻滤波器和带通滤波器

这种类型的滤波器很容易使用前两节的概念来构建。表 4.6 给出了理想、布特沃斯和高斯带阻滤波器的表达式，其中$D(u, v)$是距频率矩形中点的距离，如式(4.8-2)给出的那样，D_0是带宽的径向中心，W是带宽。图 4.63(a)以图像形式显示了一个高斯带阻滤波器，其中黑色是 0，白色是 1。

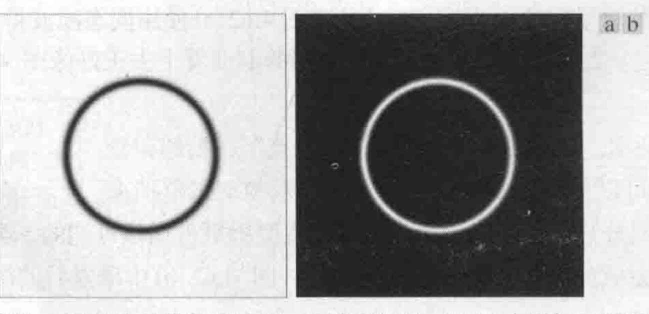
表 4.6 带阻滤波器。W是带宽，D是$D(u, v)$距滤波器中心的距离，D_0是截止频率，n是布特沃斯滤波器的阶数。为简化表中符号的表示，我们用D来代替$D(u, v)$

理 想	布特沃斯	高 斯
$H(u,v)=\begin{cases}0, & 若\ D_0-\dfrac{W}{2}\leqslant D\leqslant D_0+\dfrac{W}{2}\\ 1, & 其他\end{cases}$	$H(u,v)=\dfrac{1}{1+\left[\dfrac{DW}{D^2-D_0^2}\right]^{2n}}$	$H(u,v)=1-\mathrm{e}^{-\left[\frac{D^2-D_0^2}{DW}\right]^2}$

用由低通滤波器得到高通滤波器的相同方法，由带阻滤波器我们得到带通滤波器：

$$H_{\mathrm{BP}}(u,v)=1-H_{\mathrm{BR}}(u,v) \tag{4.10-1}$$

图4.63(b)以图像形式显示了一个高斯带通滤波器。



316~317

图 4.63 (a)带阻高斯滤波器；(b)相应的带通滤波器。为清晰起见，图(a)中添加了细黑边，它不是数据的一部分

4.10.2 陷波滤波器

陷波滤波器是更有用的选择性滤波器。陷波滤波器拒绝(或通过)事先定义的关于频率矩形中心的一个邻域的频率。零相移滤波器必须是关于原点对称的，因此，一个中心位于(u_0, v_0)的陷波在位置$(-u_0, -v_0)$必须有一个对应的陷波。陷波带阻滤波器可以用中心已被平移到陷波滤波器中心的高通滤波器的乘积来构造。一般形式为

$$H_{\text{NR}}(u,v)=\prod_{k=1}^{Q}H_k(u,v)H_{-k}(u,v) \tag{4.10-2}$$

式中，$H_k(u, v)$和$H_{-k}(u, v)$是高通滤波器，它们的中心分别是(u_k, v_k)和$(-u_k, -v_k)$。这些中心是根据频率矩形的中心$(M/2, N/2)$确定的。对于每个滤波器，距离的计算由下式执行：

$$D_k(u,v)=\left[(u-M/2-u_k)^2+(v-N/2-v_k)^2\right]^{1/2} \tag{4.10-3}$$

和

$$D_{-k}(u,v)=\left[(u-M/2+u_k)^2+(v-N/2+v_k)^2\right]^{1/2} \tag{4.10-4}$$

例如，下面是一个 n 阶布特沃斯陷波带阻滤波器，它包含三个陷波对：

$$H_{\text{NR}}(u,v)=\prod_{k=1}^{3}\left[\frac{1}{1+\left[D_{0k}/D_k(u,v)\right]^{2n}}\right]\left[\frac{1}{1+\left[D_{0k}/D_{-k}(u,v)\right]^{2n}}\right] \tag{4.10-5}$$

其中，D_k和D_{-k}由式(4.10-3)和式(4.10-4)给出。常数D_{0k}对每个陷波对都是相同的，但对于不同的陷波对，它可以不同。其他陷波带阻滤波器可用相同的方法构建，具体取决于所选的高通滤波器。正如早期讨论的滤波器那样，陷波带通滤波器可使用下式由陷波带阻滤波器得到：

$$H_{\text{NP}}(u,v)=1-H_{\text{NR}}(u,v) \tag{4.10-6}$$

如接下来的三个例子所示，陷波滤波的主要应用之一是，选择性地修改 DFT 的局部区域。这种类型的典型处理是交互完成的，它直接对 DFT 处理，而不需要填充。交互处理实际的 DFT(相对于必须从填充“平移”到实际频率值来说)的优点是，不会出现因在滤波处理中未使用填充而导致的任何缠绕错误。此外，如 5.4.4 节中将说明的那样，陷波滤波技术比这里讨论的以无填充的 DFT 为基础的技术更有用。要了解随填充函数变化的 DFT 的概念，见习题 4.22。

例 4.23 使用陷波滤波减少莫尔(波纹)模式。

图4.64(a)是来自图4.21的扫描报纸的图像，它显示了突出的莫尔模式，图4.64(b)是它的谱。由表4.3可知，纯正弦周期函数的傅里叶变换是一对共轭对称的冲激。在图4.64(b)中，对称“类冲激”是莫尔模式的近似周期性的结果。我们可以利用陷波滤波来减小这些脉冲。

图4.64(c)显示了使用一个布特沃斯陷波带阻滤波器乘以图4.64(a)的DFT的结果，对所有陷波对，该滤波器的参数为$D_0 = 3$和$n = 4$。半径值选择为(用频谱的视觉检测)包围脉冲的全部能量，将n值选择为给出陷波的适度过渡形状。陷波的中心位置由其频谱交互确定。图4.64(d)显示了使用该滤波器并遵照 4.7.3 节给出的过程得到的结果。考虑到低分辨率和原图像的退化，改进是明显的。

316 ~ 319

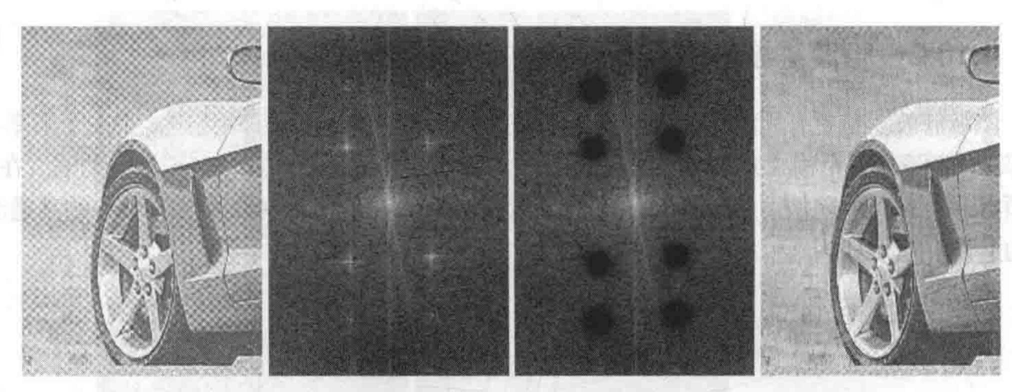


图 4.64 (a)显示了莫尔模式的取样过的报纸图像；(b)该图像的谱；(c)布特沃斯陷波带阻滤波器乘以傅里叶变换后的结果；(d)滤波后的图像

例 4.24 使用陷波滤波增强“卡西尼”号拍摄的土星图像。

图4.65(a)显示了部分环绕土星的土星环的图像。“卡西尼”号飞船首次进入该行星的轨道时，拍摄了这幅图像。垂直的正弦模式是在对图像数字化之前，由叠加到摄影机视频信号上的交流信号造成的。这是一个人们未曾料到的问题，它污染了来自某些任务的图像。所幸的是，这些类型的干扰很容易使用后处理方法来校正。一种方法就是使用陷波滤波。

图4.65(b)显示了 DFT 谱。仔细分析垂直轴会发现一系列小能量脉冲，这些脉冲对应于近似正弦曲线的干扰。一种简单的方法是使用窄陷波矩形滤波器，它从最低频脉冲开始，并扩展到垂直轴的剩余部分。图4.65(c)显示了这样一个滤波器(白色代表 1，黑色代表 0)。图4.65(d)显示了使用该滤波器对污染图像滤波后的结果。该结果与原图像相比有较大的改进。

我们用陷波带通滤波器［见图 4.66(a)］隔离了垂直轴上的频率。然后，如图 4.66(b)所示，使用这些频率的 IDFT 得到了空间干扰模式本身。

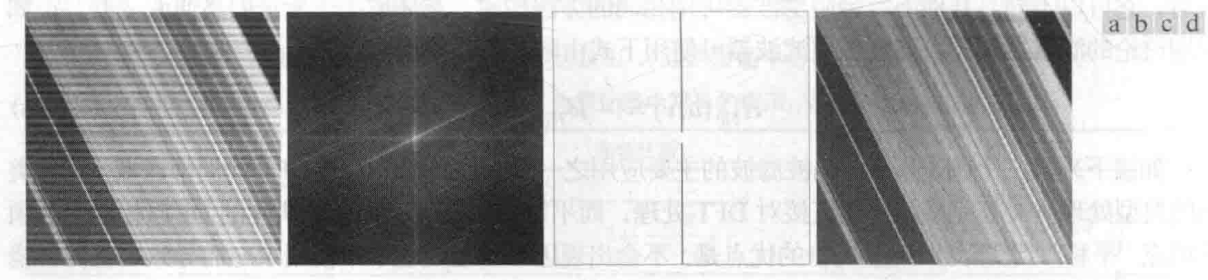


图 4.65 (a)显示了近似周期性干扰的土星环图像，图像大小为 674 × 674 像素；(b)谱：垂直轴上靠近原点的对应干扰模式的能量脉冲；(c)一个垂直陷波带阻滤波器；(d)滤波后的结果。图(c)中的细黑边是为清楚起见而添加的，它不是数据的一部分(原图像由 NASA/JPL 的 Robert A. West 博士提供)

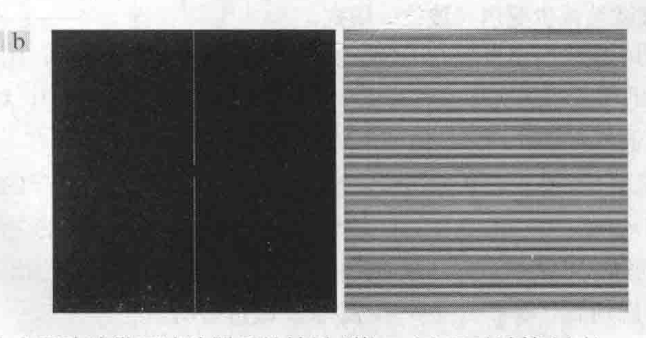


图 4.66 (a)对图 4.65(a)应用陷波带通滤波器后的结果(谱)；(b)通过计算(a)的 IDFT 而得到的空间模式

319~320

4.11 实现

迄今为止，我们的注意力集中在频率域滤波的理论概念和实例上。现在，需要清楚的一件事情是，图像处理领域的计算要求并不是微不足道的。因此，推导一些基本方法很重要，这些基本方法可简化傅里叶变换的计算，并加快计算速度。本节将处理这些问题。

4.11.1 二维 DFT 的可分性

如表 4.2 提到的那样，二维 DFT 可分成一维变换。我们可以把式(4.5-15)写为

$$F(u,v)=\sum_{x=0}^{M-1}\mathrm{e}^{-\mathrm{j}2\pi ux/M}\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi vy/N}=\sum_{x=0}^{M-1}F(x,v)\mathrm{e}^{-\mathrm{j}2\pi ux/M} \tag{4.11-1}$$

式中，

$$F(x,v)=\sum_{y=0}^{N-1}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi vy/N} \tag{4.11-2}$$

对于 x 的每个值和 $v=0,1,2,\cdots,N-1$，我们看到，$F(x,v)$就是$f(x,y)$的行的一维 DFT。在式(4.11-2)中，X 从 0 到 $M-1$ 变化，我们对$f(x,y)$的所有行计算一组一维 DFT。类似地，在式(4.11-1)中的计算是 $F(x,v)$ 的列的一维变换。

这样，我们就得出结论：$f(x,y)$的二维 DFT 可通过计算 $f(x,y)$的每一行的一维变换，然后沿计算结果的每一列计算一维变换来得到。这是一个重要的简化，因为我们必须一次只处理一个变量。类似地，我们可用一维 IDFT 来计算二维 IDFT。然而，如下节所示，我们可以用计算 DFT 所设计的算法来计算 IDFT。

> 我们可以一维列变换和一维行变换的形式来表示式(4.11-1)和式(4.11-2)。最终结果是相同的。

4.11.2 用 DFT 算法计算 IDFT

取式(4.5-16)两边的复共轭，并乘以 MN 得到

$$MNf^*(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}F^*(u,v)\mathrm{e}^{-\mathrm{j}2\pi(ux/M+vy/N)} \tag{4.11-3}$$

但是，我们看出该结果右边的形式类似于$F^*(u,v)$的DFT。因此，式(4.11-3)指出，如果把 $F^*(u,v)$代入计算二维傅里叶正变换所设计的算法中，结果将是 $MN\,f^*(x,y)$。取复共轭并将该结果乘以 $1/MN$，将得到 $f(x,y)$，它是 $F(u,v)$的反变换。

> 在这种推导中乘以 MN 的假设是针对式(4.5-15)和式(4.5-16)的形式的。如果常数在正反变换间的分布不同，则需要不同的常数乘数方案。

由(前节介绍的)基于逐次通过一维变换的二维正 DFT 算法来计算二维反变换的方法，是导致混乱的一种常见原因，因为这涉及复共轭和乘以一个常数的计算，而在一维算法中这两种计算都不会涉及。应记住的一个关键概念是，我们只需简单地将 $F^*(u,v)$输入已有的正向算法中。结果将是 $MNf^*(x,y)$。要得到$f(x,y)$这一结果，我们只需取它的复共轭并乘以一个常数 $1/MN$。当然，当$f(x,y)$是实数时，典型的情况是 $f^*(x,y)=f(x,y)$。

4.11.3 快速傅里叶变换(FFT)

如果我们必须直接实现式(4.5-15)和式(4.51-16)，那么在频率域工作并不实际。这些公式的大规模运算要求约$(MN)^2$次求和及加法操作。对于中等大小的图像(譬如 1024×1024 像素)，这意味着仅进行

300 ~ 321

一次 DFT 就需要万亿次量级的乘法和加法，还不包括指数运算，指数运算可被计算一次并存储在查找表中。即使是对超级计算机来说，这也是一种挑战。在人们未发现快速傅里叶变换(FFT)前，可以说本章的内容基本不具有实践价值，快速傅里叶变换可将乘法和加法的次数降到 $MN\ \log_2 MN$。由 FFT 提供的计算量的减少的确令人印象深刻。例如，计算大小为 1024×1024 的图像的二维 FFT，只需要约 2000 万次乘法和加法运算，与前面提到万亿次乘法与加法运算相比，降低非常明显。

虽然 FFT 是信号处理文献中广泛涵盖的主题，但这一主题在本章中也非常重要，如果我们不提供至少介绍 FFT 是如何工作的说明，那么本章也将不完整。我们选择完成这一目的的算法称为*逐次加倍方法*，它是导致整个产业诞生的原始算法。这种特殊算法假定取样数是 2 的整数次幂，但这不是其他方法的通用要求(Brigham[1988])。由 4.11.1 节可知，二维 DFT 可以使用逐次一维变换的方法来执行，因此我们仅需关注一个变量的 FFT。

在推导 FFT 时，习惯上将式(4.4-6)表达为如下形式：

$$F(u)=\sum_{x=0}^{M-1} f(x)W_M^{ux} \tag{4.11-4}$$

式中，$u=0,1,\cdots,M-1$，其中，

$$W_M=\mathrm{e}^{-\mathrm{j}2\pi/M} \tag{4.11-5}$$

且假设 M 具有如下形式：

$$M=2^n \tag{4.11-6}$$

其中 N 是一个正整数。因此，M 可表示为

$$M=2K \tag{4.11-7}$$

K 也是一个正整数。将式(4.11-7)代入式(4.11-4)得到

$$F(u)=\sum_{x=0}^{2K-1} f(x)W_{2K}^{ux}=\sum_{x=0}^{K-1} f(2x)W_{2K}^{u(2x)}+\sum_{x=0}^{K-1} f(2x+1)W_{2K}^{u(2x+1)} \tag{4.11-8}$$

然而，使用式(4.11-5)可以证明 $W_{2K}^{2ux}=W_K^{ux}$，因此式(4.11-8)可以表示为

$$F(u)=\sum_{x=0}^{K-1} f(2x)W_K^{ux}+\sum_{x=0}^{K-1} f(2x+1)W_K^{ux}W_{2K}^{u} \tag{4.11-9}$$

定义

$$F_{\text{even}}(u)=\sum_{x=0}^{K-1} f(2x)W_K^{ux} \tag{4.11-10}$$

式中，$u=0,1,2,\cdots,K-1$，和

$$F_{\text{odd}}(u)=\sum_{x=0}^{K-1} f(2x+1)W_K^{ux} \tag{4.11-11}$$

式中，$u=0,1,2,\cdots,K-1$，将式(4.11-9)简化为

$$F(u)=F_{\text{even}}(u)+F_{\text{odd}}(u)W_{2K}^{u} \tag{4.11-12}$$


321 ~ 323

此外，因为 $W_M^{u+M}=W_M^u$ 和 $W_{2M}^{u+M}=-W_{2M}^u$，式(4.11-10)到式(4.11-12)给出

$$F(u+K)=F_{\text{even}}(u)-F_{\text{odd}}(u)W_{2K}^{u} \tag{4.11-13}$$

分析式(4.11-10)到式(4.11-13)可以发现这些表达式的一些有趣特性。如式(4.11-12)和式(4.11-13)指出的那样，一个 M 点变换可以通过把原始表达式分解为两部分来计算。计算 $F(u)$的前一半要求计算式(4.11-10)和式(4.11-11)中给出的两个$(M/2)$点变换。然后，将结果即 $F_{\text{even}}(u)$ 和 $F_{\text{odd}}(u)$ 的值代入式(4.11-12)得到 $F(u), u = 0, 1, 2, \cdots, (M/2-1)$。再后，可由式(4.11-13)直接计算 $F(u)$的另一半，而不需要额外的变换计算。

为检验该步骤的计算量，令 $m(n)$和 $a(n)$分别表示实现算法所要求的复数乘法次数和加法次数。如前所述，样本数为 2^n，n 为正整数。先假设 $n=1$。两点变换要求计算 $F(0)$；然后 $F(1)$由式(4.11-13)得到。为得到 $F(0)$，要求计算 $F_{\text{even}}(0)$和 $F_{\text{odd}}(0)$。此例中$K=1$ 且式(4.11-10)和式(4.11-11)是单点变换。然而，由于单个样本点的 DFT 是该样本自身，故得到 $F_{\text{even}}(0)$和 $F_{\text{odd}}(0)$不需要任何乘法和加法运算。$F_{\text{odd}}(0)$与 W_2^0 的一次相乘和一次加法，即可由式(4.11-12)得到 $F(0)$。然后，由式(4.11-13)得到 $F(1)$，又用了一次加法(减法可以认为与加法相同)。因为已计算过 $F_{\text{odd}}(0)W_2^0$，故两点变换所需要的总运算次数为 $m(1)=1$ 次乘法和 $a(1)=2$ 次加法。

n 的下一个允许值为 2。根据前面的推导，一个 4 点变换可分成两部分。$F(u)$的前半部分要求计算两个两点变换，如 $K=2$ 时式(4.11-10)和式(4.11-11)给出的那样。如前一段中解释的那样，一个两点变换要求 $m(1)$次乘法和 $a(1)$次加法，因此这两个等式的计算总共要求 $2m(1)$次乘法和 $2a(1)$次加法。要从式(4.11-12)得到 $F(0)$和 $F(1)$，还需要两次乘法和加法。因为 $F_{\text{odd}}(u)W_{2K}^u$ 在 $u=\{0, 1\}$时已被计算过，所以要得到 $F(2)$和 $F(3)$又要进行两次加法计算。这时，总运算次数是 $m(2)=2m(1)+2$ 次乘法和 $a(2)=2a(1)+4$ 次加法。

当 $n=3$ 时，计算 $F_{\text{even}}(u)$和 $F_{\text{odd}}(u)$时会用到两个 4 点变换。它们需要 $2m(2)$次乘法和 $2a(2)$次加法。完成全部变换还要 4 次乘法和 8 次加法运算。总运算次数为 $m(3)=2m(2)+4$ 次乘法和 $a(3)=2a(2)+8$ 次加法。

对任意正整数 n 计算这些参量，推导出完成 FFT 所需乘法和加法次数的递归表达式为

$$m(n) = 2m(n-1) + 2^{n-1} \quad n \geq 1 \tag{4.11-14}$$

和

$$a(n) = 2a(n-1) + 2^n \quad n \geq 1 \tag{4.11-15}$$

其中，$m(0)=0$ 且 $a(0)=0$，因为单点变换无须任何加法和乘法运算。

式(4.11-10)至式(4.11-13)的实现构成了逐次加倍的 FFT 算法。这一名称来自如下计算方法：对于任意的 2 的整数次幂的 M，一个两点变换来自两个单点变换，一个 4 点变换来自两个两点变换，以此类推，直到 M 为 2 的整数次幂的情形。作为练习(见习题 4.41)，请读者证明

$$m(n) = \frac{1}{2} M \log_2 M \tag{4.11-16}$$

和

$$a(n) = M \log_2 M \tag{4.11-17}$$

与直接计算一维 DFT 相比，FFT 的计算优势定义为

$$c(M) = \frac{M^2}{M \log_2 M} = \frac{M}{\log_2 M} \tag{4.11-18}$$

因为假定有 $M=2^n$，所以可将式(4.11-18)写为关于 n 的形式：

323~324

$$c(n)=\frac{2^n}{n} \tag{4.11-19}$$

图 4.67 显示了该函数的曲线。作为 n 的函数，迅速增加的计算优势是明显的。例如，当 $n=15$(32 768 点)时，FFT 与 DFT 相比有接近 2200:1 的优势。这样，我们可以预料在同一台机器上，FFT 的计算速度会比 DFT 的计算速度快近 2200 倍。

涵盖 FFT 细节的优秀文献很多，因此此处不再深入地讨论这一主题(例如，可参阅 Brigham[1988])。实际上，所有综合性的信号和图像处理软件包都已包含 FFT 的实现，也能够处理点数不为 2 的整数次幂的情况(以牺牲计算效率为代价)。免费的 FFT 程序也可以从 Internet 上得到。

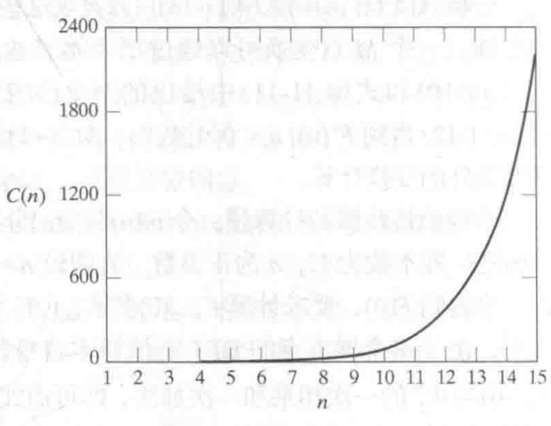


图 4.67 FFT 相对于直接实现一维 DFT 的计算优势。注意，该优势以 n 的函数快速增长

4.11.4 关于滤波器设计的一些注释

本章中讨论的滤波方法严格基于基础知识，重点是尽可能清楚地解释频率域滤波的效果。除了这样做之外，我们知道没有更好的方法来处理滤波问题。我们可将这种处理方式视为“原型化”一个滤波器的基础。换句话说，若给出一个我们需要找到一种滤波器来解决的问题，那么频率域方法是一种试验、快速及完全控制滤波器所有参数的理想工具。

一旦找到适用于某个应用的滤波器，通常会使用固件和/或硬件直接在空间域实现该滤波器。该主题超出了本书的范围。Petrou and Bosdogianni[1999]给出了二维频率域滤波器和相应数字滤波器之间的良好联系。关于二维数字滤波器的设计，请参阅 Lu and Antoniou[1992]。

小结

本章循序渐进地介绍了取样、傅里叶变换和频率域滤波。某些概念，如取样定理，如果没有频率域内容的解释，几乎没有意义。混淆效应同样如此。因此，前面几节中介绍的内容，是理解数字图像处理基本原理的坚实基础。我们特别注意用基本原理来开始内容的阐述，因此许多具有现代数学背景的读者不仅可以尽可能地吸收这些材料，而且还可应用它们。

本章的第二个目的是阐述离散傅里叶变换及其在频率域滤波中的应用。为实现这一目标，我们介绍了卷积定理。这一结果是线性系统的基础，也是第 5 章中将要介绍的复原技术的基础。我们讨论过的滤波器类型是在实践中发现的代表性滤波器。然而，对于这些滤波器，关键是说明用公式形式来表达并在频率域实现有多么简单。尽管一种典型解决方案的最终实现还是以空间滤波器为基础的，但工作于频率域所获得的知识对于选择空间滤波器来说大有裨益。

虽然本章中的多数滤波例子是在图像增强领域，但这些过程本身是通用的，并且广泛用于后续章节。

参考文献

324 ~ 326

4.1 节所讲内容的附加读物见 Hubbard[1998]。Bracewell[2000, 1995]介绍了连续傅里叶变换及其二维图像处理。除 Lim[1990]，Castleman[1996]，Petrou and Bosdogianni[1999]，Brigham[1988]和 Smith[2003]，这两本图书为 4.2 节到 4.6 节的大部分内容提供了背景知识。关于莫尔模式的早期工作综述，请参阅 Oster

and Nishijima[1963]。Creath and Wyant[1992]讨论了 30 年后该领域的技术状况。4.5 节讨论的取样、混淆和图像重建问题也是计算机图形学中感兴趣的主题，正如 Shirley[2002]举例说明的那样。

4.7 节到 4.11 节内容的附加背景材料，可参阅 Castleman[1996]，Pratt[2001]和 Hall[1979]。要了解“卡西尼”号太空飞船(见 4.10.2 节)中成像传感器的更多内容，请参阅 Porco, West et al.[2004]。滤波器实现的有效处理问题(如振铃)一直是令人感兴趣的主题，详见 Bakir and Reeves[2000]。关于钝化模板和高频强调滤波，见 Schowengerdt[1983]。同态滤波(见 4.9.5 节)的内容基于 Stockham[1972]，也可参见 Oppenheim and Schafer[1975]和 Pitas and Venetsanopoulos[1990]。Brinkman et al.[1998]结合钝化模板和同态滤波，增强了核磁共振图像。

如 4.1.1 节注释的那样，作为信号处理基础工具的 DFT 的普及，快速傅里叶变换的发现(4.11.3 节)是一个重要的里程碑。4.11.3 节中介绍的 FFT 基于 Cooley and Tuckey[1965]和 Brigham[1988]，他们也讨论了 FFT 的实现，包括非基 2 的情况。快速傅里叶变换的公式表达通常认为由 Cooley and Tukey[1965]创造。但 FFT 的有趣历史在此值得一提。作为对 Cooley and Tuckey[1965]的响应，Rudnick[1966]提出了一种类似的技术，这种技术的运算量同样与 $N\log_2 N$ 成正比，并且基于 Danielson and Lanczos[1942]。这些作者依次参考了 Runge[1903, 1905]。稍后的两篇论文和 Runge and König[1924]中，包含了现代快速傅里叶变换算法的基本计算优势。Yates[1937]，Stumpff[1939]，Good[1958]和 Thomas[1963]也发表过类似的技术。1965 年，在 Cooley-Tukey 的论文发表前，Cooley, Lewis and Welch[1967a]对快速傅里叶变换进行了历史性的概括，并对各种结果进行了有趣的对比。

4.11.3 节介绍的 FFT 算法来自 Cooley and Tukey[1965]，补充读物见 Brigham[1988]和 Smith[2003]。本章中讨论的基于频率域的数字滤波器的设计(见 4.11.4 节)，请参阅 Lu and Antoniou [1992]和 Petrou and Bosdogianni[1999]。4.7 节到 4.11 节讨论的许多软件实现方法，见 Gonzalez, Woods and Eddins[2004]。

标有星号习题的详细解答可在本书的网站上找到。该网站上还包含有基于本章内容而建议的项目。

习题

4.1 重复例 4.1，但使用函数 $f(t)=2A(-W/4\leqslant t\leqslant W/4)$和 $f(t)=0$，对于其他所有的 t 值。对你的结果和例子中的结果之间的任何不同，解释原因。

★**4.2** 证明式(4.4-2)中的 $\tilde{F}(\mu)$ 在两个方向上是无限周期的，周期为 $1/\Delta T$ 。

★**4.3** 可以证明(Bracewell[2000]) $1\Leftrightarrow\delta(\mu)$ 和 $\delta(t)\Leftrightarrow 1$ 。使用前一个性质和表 4.3 中的平移性质，证明连续函数 $f(t)=\cos(2\pi nt)$ 的傅里叶变换是 $F(\mu)=(1/2)[\delta(\mu+n)+\delta(\mu-n)]$，其中 n 是一个实数。

4.4 考虑连续函数 $f(t)=\cos(2\pi nt)$ 。

★(a) $f(t)$的周期是什么？

★(b) $f(t)$的频率是什么？

$f(t)$的傅里叶变换 $F(\mu)$ 是实的(见习题4.3)，因为取样数据的傅里叶变换由 $F(\mu)$ 的周期副本组成，取样数据的傅里叶变换 $\tilde{F}(\mu)$ 也将是实的。画出一个类似于图 4.6 的图表，并根据所绘图表(假定取样开始于 $t=0$)回答下面的问题。

★(c) 如果 $f(t)$以高于奈奎斯特率的取样率来取样，取样函数将是什么？取样后的函数及其傅里叶变换在图中看上去是什么样子？

(d) 如果 $f(t)$以低于奈奎斯特率的取样率取样，那么取样后的函数通常看上去像什么？

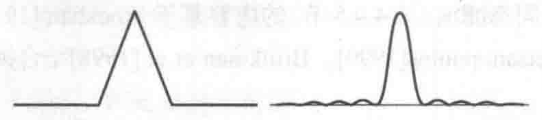
(e) 如果 $f(t)$以奈奎斯特率取样，且在 $t=0, \Delta T, 2\Delta T, \cdots$ 处取样，取样后的函数看上去像什么？

326 ~ 327

★**4.5** 如式(4.2-21)和式(4.2-22)给出的那样，证明连续函数的一维卷积定理的正确性。

4.6 完成由式(4.3-11)和式(4.3-12)给出的步骤。

4.7 如下图所示，一个“帐篷形”函数(左侧)的傅里叶变换是一个平方正弦函数(右侧)。提出一个参量来表明一个“帐篷形”函数的傅里叶变换可以由一个盒状函数的傅里叶变换得到。(提示：“帐篷形”函数本身可以由两个相等的盒状函数进行卷积操作生成。)



4.8 (a) 证明式(4.4-4)和式(4.4-5)组成傅里叶变换对。

★(b) 对式(4.4-6)和式(4.4-7)重复(a)。在该习题的两部分中，需要用到如下的指数正交性质：

$$\sum_{x=0}^{M-1} \mathrm{e}^{\mathrm{j}2\pi rx/M}\mathrm{e}^{-\mathrm{j}2\pi ux/M} = \begin{cases} M, & r=u \\ 0, & \text{其他} \end{cases}$$

4.9 证明式(4.4-8)和式(4.4-9)的正确性。

★**4.10** 证明一个变量的离散卷积定理的正确性［见式(4.2-21)、式(4.2-22)和式(4.2-10)］。证明时需要用到平移性质 $f(x)\mathrm{e}^{\mathrm{j}2\pi u_0 x/M} \Leftrightarrow F(u-u_0)$ 和反平移性质 $f(x-x_0) \Leftrightarrow F(u)\mathrm{e}^{-\mathrm{j}2\pi x_0 u/M}$。

★**4.11** 写出二维连续卷积的表达式。

4.12 考虑一幅棋盘图像，其中每个方格的大小为 0.5 mm × 0.5 mm。假定图像在两个坐标方向上无限扩展，为避免混淆，问最小取样率是多少(样本数/mm)？

4.13 由 4.5.4 节中的讨论可知，收缩图像会导致混淆现象。放大图像也会导致混淆现象吗？请解释。

★**4.14** 证明一维连续和离散傅里叶变换都是线性操作(见 2.6.2 节关于线性的定义)。

4.15 假定读者有一个用于计算二维 DFT 对的“已封装”程序，但不知道 1/*MN* 项是包含在两个公式中的一个公式中，还是分成两个常数 $1/\sqrt{MN}$ 放在正变换和反变换前面。如果程序文档中没有这种信息，请问如何才能找到包含该项(或这些项)的位置。

4.16 证明连续和离散二维傅里叶变换都是平移和旋转不变的。

4.17 由习题 4.3 可以推出 $1 \Leftrightarrow \delta(\mu,\nu)$ 和 $\delta(t,z) \Leftrightarrow 1$。使用前一个性质和表 4.3 中的平移性质，证明连续函数 $f(t,z)=A\cos(2\pi\mu_0 t + 2\pi\nu_0 z)$ 的傅里叶变换是

$$F(\mu,\nu) = \frac{1}{2}\left[\delta(\mu+\mu_0,\nu+\nu_0) + \delta(\mu-\mu_0,\nu-\nu_0)\right]$$

4.18 证明离散函数 $f(x,y)=1$ 的 DFT 是

$$\Im\{1\} = \delta(u,v) = \begin{cases} 1, & \text{若 } u=v=0 \\ 0, & \text{其他} \end{cases}$$

4.19 证明离散函数 $f(x,y)=\cos(2\pi u_0 x + 2\pi v_0 y)$ 的 DFT 是

$$F(u,v) = \frac{1}{2}\left[\delta(u+Mu_0, v+Nv_0) + \delta(u-Mu_0, v-Nv_0)\right]$$

4.20 下列问题与表 4.1 中的性质有关。

★(a) 证明性质 1 的正确性。

★(b) 证明性质 3 的正确性。

(c) 证明性质 6 的正确性。

327 ~ 328




★(d) 证明性质 7 的正确性。

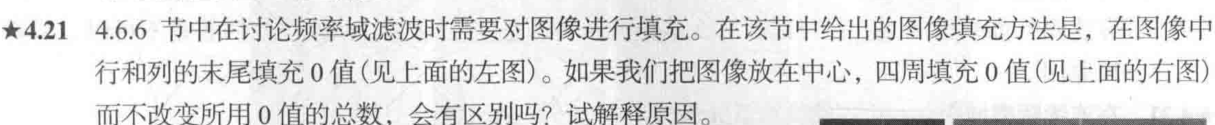
(e) 证明性质 9 的正确性。

(f) 证明性质 10 的正确性。

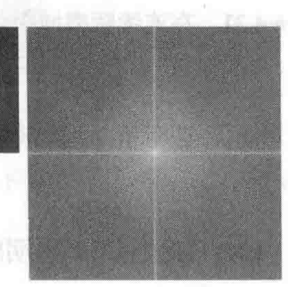
★(g) 证明性质 11 的正确性。

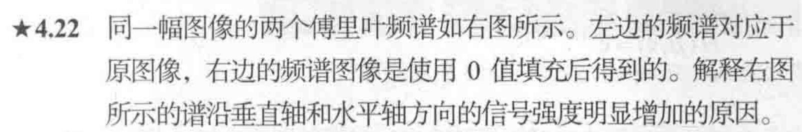
(h) 证明性质 12 的正确性。

(i) 证明性质 13 的正确性。



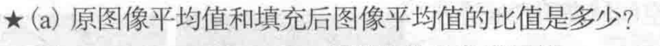
★4.21 4.6.6 节中在讨论频率域滤波时需要对图像进行填充。在该节中给出的图像填充方法是，在图像中行和列的末尾填充 0 值(见上面的左图)。如果我们把图像放在中心，四周填充 0 值(见上面的右图)而不改变所用 0 值的总数，会有区别吗？试解释原因。





★4.22 同一幅图像的两个傅里叶频谱如右图所示。左边的频谱对应于原图像，右边的频谱图像是使用 0 值填充后得到的。解释右图所示的谱沿垂直轴和水平轴方向的信号强度明显增加的原因。

4.23 由表 4.2 可知 DFT 的直流项 $F(0,0)$ 与其对应的空间图像的平均值成正比。假定图像尺寸是 $M \times N$。假如对图像进行零填充后，图像的尺寸为 $P \times Q$，其中 P 和 Q 分别由式(4.6-31)和式(4.6-32)给出。令 $F_P(0,0)$ 代表填充后的函数的 DFT 的直流项。



★(a) 原图像平均值和填充后图像平均值的比值是多少？

(b) $F_p(0,0)=F(0,0)$ 吗？假设从数学角度回答。

4.24 证明表 4.2 中的周期性质(性质 8)。

4.25 下列问题与表 4.3 中的性质有关。

★(a) 证明一维情况下离散卷积定理(性质 8)的正确性。

(b) 对于二维情况，重复(a)。

(c) 证明性质 9 的正确性。

(d) 证明性质 13 的正确性。

(注意：习题 4.18、习题 4.19 和习题 4.31 也与表 4.3 有关。)

4.26 (a) 证明连续变量 t 和 z 的连续函数 $f(t,z)$ 的拉普拉斯变换满足下列傅里叶变换对［拉普拉斯变换的定义见式(3.6-3)］：

$$\nabla^2 f(t,z) \Leftrightarrow -4\pi^2(\mu^2+\nu^2)F(\mu,\nu)$$

［提示：研究表 4.3 中的性质 12 并参阅习题 4.25(d)。］

★(b) 前面闭合形式的表达式仅适用于连续变量。然而，使用 $M \times N$ 滤波器

$$H(u,v) = -4\pi^2(u^2+v^2)$$

它可能是在离散频率域实现拉普拉斯的基础，$u=0,1,2,\cdots,M-1$，$v=0,1,2,\cdots,N-1$。解释你怎样实现这个滤波器。

(c) 如你在例 4.20 中看到的那样，频率域的拉普拉斯结果类似于使用中心系数为−8 的空间模板的结果。请说明频率域拉普拉斯的结果与使用中心系数为−4 的空间模板的结果不同的原因。见 3.6.2 节关于空间域的拉普拉斯算子。

★4.27 考虑大小为 5×5 的空间模板，它平均与点 (x,y) 最靠近的 12 个邻点，但平均中排除该点本身。

(a) 在频率域找出与其等价的滤波器 $H(u,v)$。

(b) 证明你的结果是一个低通滤波器。

328~330

4.28 基于式(3.6-4)，近似二维离散微分的一种方法是计算$f(x+1, y) + f(x-1, y) - 2f(x, y)$和$f(x, y+1) + f(x, y-1)$的差。

(a) 在频率域找出与其等价的滤波器$H(u, v)$。

(b) 证明你的结果是一个高通滤波器。

4.29 找出一个等价的滤波器$H(u, v)$，它在频率域实现使用图3.37(a)中的拉普拉斯模板执行的空间操作。

★**4.30** 你能想出一种使用傅里叶变换计算(或分部计算)用于图像差分的梯度幅度［见式(3.6-11)］的方法吗？如果你的回答是可以，那么请给出一种方法去实现它。如果你的回答是不可以，请解释原因。

★**4.31** 在连续频率域中，一个连续高斯低通滤波器有如下传递函数：

$$H(\mu, v) = \mathrm{e}^{-(\mu^2+v^2)}$$

证明相应的空间域滤波器是

$$h(t, z) = \pi \mathrm{e}^{-2\pi^2(t^2+z^2)}$$

4.32 如式(4.9-1)说明的那样，从低通滤波器的传递函数得到高通滤波器的传递函数H_{HP}是可能的：

$$H_{\mathrm{HP}} = 1 - H_{\mathrm{LP}}$$

使用习题4.31中给出的信息，回答空间域高斯高通滤波器是什么形式。



4.33 考虑右侧所示的图像。右图是通过如下步骤得到的：(a)用$(-1)^{x+y}$乘以左图；(b)计算其DFT；(c)取该变换的复共轭；(d)计算反DFT；(e)用$(-1)^{x+y}$乘以结果的实部。(从数学上)解释为什么右图会出现该现象。

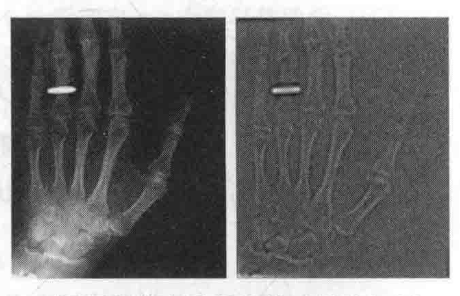
4.34 图4.41(b)的水平轴上近似周期性的亮点的来源是什么？

★**4.35** 图4.53中的每个滤波器在其中心处都有一个很强的尖刺，解释这些尖刺的来源。

4.36 考虑下面所示的图像。右图是对左图用高斯低通滤波器进行低通滤波，然后用高斯高通滤波器对结果再进行高通滤波得到的。图像的大小为420×344，两个滤波器均使用了$D_0 = 25$。

(a) 解释右图中戒指的中心部分明亮且实心的原因，考虑滤波后图像的支配特性是物体(如手指、腕骨)外边界上的边缘及这些边缘之间的暗区域。换句话说，你并不希望高通滤波器将戒指内部的恒定区域渲染为暗色，因为高通滤波消除了直流项。

(b) 如果颠倒滤波处理的顺序，你认为结果会有区别吗？



(原图像由密歇根大学医学院解剖学分部的Thomas R. Gest博士提供)

4.37 给出一幅大小为$M \times N$的图像，要求做一个实验，实验使用截止频率为D_0的高斯低通滤波器重复对该图像进行低通滤波。可以忽略计算上的舍入误差。令$c_{\min}$是实验所用机器可表示的最小正数。

★(a) 令K表示该滤波器使用的次数。在进行实验前，你能预测K为足够大的值时的结果(图像)将是什么吗？如果能，结果是什么？

(b) 推导保证预测结果的最小K值的表达式。

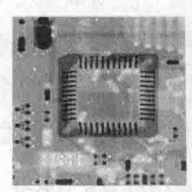
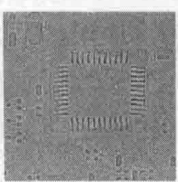
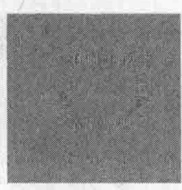
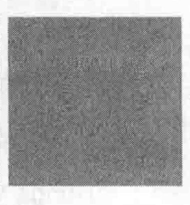
330～331

4.38 考虑下面所示的图像序列。最左侧的图像是商用印制电路板的X射线图像的一部分。该图像右侧的图像分别是使用一个$D_0 = 30$的高斯高通滤波器进行1次、10次和100次滤波后的结果。图像的

大小为 330 × 334 像素，每个像素由 8 比特灰度表示。为便于显示，图像已进行了缩放，但这对本习题没有影响。

(a) 从这几幅图像可以看出，经过有限次数的滤波后，图像将不再发生变化。请说明实际是否如此。可以忽略计算舍入误差。令 $c_{\min}$ 表示完成此实验的机器可表示的最小正数。

(b) 如果在(a)中确定有限次迭代后变化将停止，求最小的迭代次数。

(原图像由 Lixi 公司的 Joseph E. Pascente 先生提供)

4.39 如图 4.59 中说明的那样，将高频强调和直方图均衡相结合是实现边缘锐化和对比度增强的有效方法。

(a) 说明这种结合方法是否与先用哪种处理有关。

(b) 如果与应用顺序有关，请给出先采用某种方法的理由。

4.40 使用一个布特沃斯高通滤波器构建一个同态滤波器，该滤波器的形状与图 4.6 中的滤波器形状相同。

★4.41 证明式(4.11-16)和式(4.11-17)的正确性。(提示：使用归纳法证明。)

4.42 假设有一组图像，这组图像是由对恒星事件分析的实验生成的。每幅图像都包含一组明亮且分散的点，这些点对应于广袤宇宙空间中的星星。问题是这些星星因为大气折射导致的重叠照射几乎不可见。如果这些图像使用一组冲激建模为一个恒定照射成分的乘积，试提出一个增强过程，它以设计为显示星星自身的图像成分的同态滤波为基础。

4.43 一种成熟的医学技术被用于检测电子显微镜生成的某类图像。为简化检测任务，技术人员决定采用数字图像增强技术，并在处理结束后，检查了一组具有代表性的图像，发现了如下问题：(1)明亮且孤立的点是不感兴趣的点；(2)清晰度不够；(3)一些图像的对比度不够；(4)平均灰度值已被改变，而正确地执行某种灰度度量的这个值应是 V。技术人员想要纠正这些问题，然后将 I_1 和 I_2 波段之间的所有灰度显示为白色，同时保持其余灰度的正常色调。请为技术人员提出达到期望目的的处理步骤。可以使用第 3 章和第 4 章的技术。

第 5 章　图像复原与重建

Things which we see are not by themselves what we see. It remains completely unknown to us what the objects may be by themselves and apart from the receptivity of our senses. We know nothing but our manner of perceiving them.

Immanuel Kant

引言

如图像增强那样，图像复原技术的主要目的是以预先确定的目标来改善图像。尽管两者的覆盖领域有重叠之处，但图像增强主要是一个主观过程，而图像复原大部分是一个客观过程。图像复原试图利用退化现象的某种先验知识来复原被退化的图像。因而，复原技术是面向退化模型的，并且采用相反的过程进行处理，以便恢复出原图像。

这种方法通常会涉及设立一个最佳准则来产生期望结果的最佳估计。相比之下，图像增强技术基本上是一个探索性过程，即根据人类视觉系统的生理特点来设计改善图像的方法。例如，对比度拉伸被认为是一种增强技术，因为它主要为观看者提供能接受的图像，而通过去模糊函数去除图像模糊则被认为是一种图像复原技术。

本章中给出的素材严格地说只是介绍性的。我们仅从给出的一幅退化数字图像的特点来考虑复原问题；因此对传感器、数字化转换器和显示退化等主题考虑得较少。尽管这些主题在图像复原应用的整个处理中非常重要，但它们超出了本章所要讨论的范围。

如第 3 章和第 4 章中讨论的那样，有些空间域复原技术已得到了很好的阐述，而另一些复原技术则更适用于频率域。例如，当退化仅是加性噪声时，空间处理就非常适用。另一方面，图像模糊这样的退化在空间域使用较小的滤波模板处理时，会很困难。在这种情况下，可以选择基于不同优化准则的频率域滤波。这些滤波器同样也考虑到了噪声的存在。如第 4 章中所讲的那样，频率域中用来解决给定应用的复原滤波器通常是生成一个数字滤波器的基础，数字滤波器的程序操作更适合于用硬件或固件来实现。

5.1 节介绍图像退化/复原过程的一个线性模型。5.2 节处理实践中经常遇到的各种噪声模型。5.3 节阐述降低图像噪声的几种空间滤波技术，即通常称为图像去噪的处理。5.4 节致力于使用频率域降噪的技术。5.5 节介绍图像退化的线性、位置不变模型，5.6 节介绍估计退化函数的方法。5.7 节至 5.10 节阐述基本的图像复原方法。本章最后（见5.11节）介绍由投影来重建图像。这一概念的主要应用是计算断层（CT），即图像处理在健康领域最为重要的一种商业应用。

333～334

5.1　图像退化/复原过程的模型

如图 5.1 所示，在本章中，退化过程被建模为一个退化函数和一个加性噪声项，对一幅输入图像 $f(x,y)$ 进行处理，产生一幅退化后的图像 $g(x,y)$。给定 $g(x,y)$ 和关于退化函数 H 的一些知识以及关于加性噪声项 $\eta(x,y)$ 的一些知识后，图像复原的目的就是获得原始图像的一个估计 $\hat{f}(x,y)$。通常，我们希望这一估计尽可能地接近原始输入图像，并且 H 和 η 的信息知道得越多，所得到的 $\hat{f}(x,y)$ 就会越接近 $f(x,y)$。在本章中使用的大部分复原方法都是以不同类型的图像复原滤波器为基础的。

如 5.5 节所示，如果 H 是一个线性的、位置不变的过程，那么空间域中的退化图像可由下式给出：

$$g(x,y) = h(x,y) \star f(x,y) + \eta(x,y) \tag{5.1-1}$$

式中，$h(x,y)$ 是退化函数的空间表示；与第 4 章相同，符号“★”表示空间卷积。由 4.6.6 节的讨论可知，空间域中的卷积等同于频率域中的乘积，因此可以把式(5.1-1)中的模型写成等价的频率域表示：

$$G(u,v) = H(u,v)F(u,v) + N(u,v) \tag{5.1-2}$$

其中的大写字母项是式(5.1-1)中相应项的傅里叶变换。这两个公式是本章中大部分复原内容的基础。

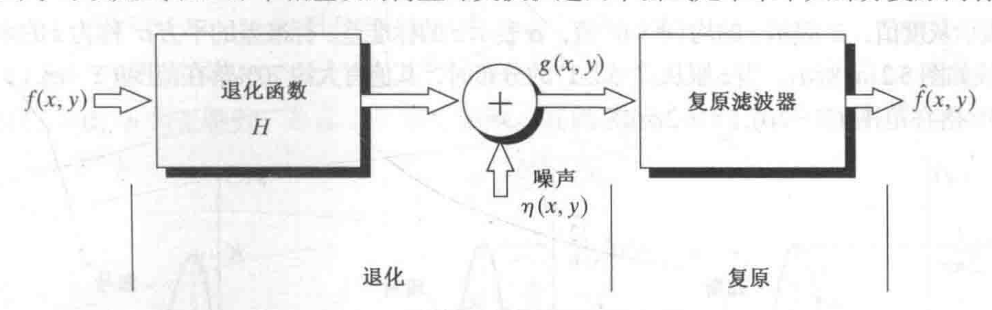


图 5.1　图像退化/复原过程的模型

在接下来的三节中，我们假设 H 是一个同一性算子，并且只处理由噪声引起的退化。从 5.6 节开始，我们考虑若干重要的图像退化函数，并考虑几个 H 和 η 同时存在的图像复原方法。

5.2　噪声模型

数字图像中，噪声主要来源于图像的获取和/或传输过程。成像传感器的性能受各种因素的影响，如图像获取过程中的环境条件和传感元器件自身的质量。例如，在使用 CCD 摄像机获取图像时，光照水平和传感器温度是影响结果图像中噪声数量的主要因素。图像在传输中被污染主要是由于传输信道中的干扰。例如，使用无线网络传输的图像可能会因为光照或其他大气因素而污染。

5.2.1　噪声的空间和频率特性

与我们的讨论相关的是定义噪声空间特性的参数，以及噪声是否与图像相关。频率特性是指傅里叶域中噪声的频率内容(即相对于电磁波谱的频率)。例如，当噪声的傅里叶谱是常量时，噪声通常称为白噪声。这个术语是从白光的物理特性派生出来的，它以相等的比例包含可见光谱中的几乎所有频率。从第 4 章的讨论中不难看出，以相同比例包含所有频率的函数的傅里叶谱是一个常量。

除了空间周期噪声(见 5.2.3 节)之外，在本章中我们假设噪声与空间坐标无关，并且噪声与图像本身不相关(即像素值与噪声成分的值之间不相关)。虽然这些假设至少在某些应用中(例如 X 射线

334 ~ 335

和核医学成像的有限量子成像就是一个很好的例子)是无效的，但处理空间相关和相关性噪声的复杂性超出了我们讨论的范围。

5.2.2　一些重要的噪声概率密度函数

基于前一节的假设，我们关心的空间噪声描述子就是图 5.1 中模型的噪声成分中灰度值的统计特性。可以认为它们是由概率密度函数(PDF)表征的随机变量。下面是在图像处理应用中最常见的 PDF。

关于概率论的简要回顾，请查阅本书的网站。

高斯噪声

在空间域和频率域中，由于高斯噪声在数学上的易处理性，故实践中常用这种噪声(也称为正态噪声)模型。事实上，这种易处理性非常方便，以至于高斯模型常常应用于在一定程度上导致最好结果的场合。

高斯随机变量 z 的 PDF 由下式给出：

$$p(z)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(z-\bar{z})^2/2\sigma^2} \tag{5.2-1}$$

式中，z 表示灰度值，$\bar{z}$ 表示 z 的均(平均)[①]值，σ 表示 z 的标准差。标准差的平方 σ^2 称为 z 的方差。高斯函数的曲线如图 5.2(a)所示。当 z 服从式(5.2-1)的分布时，其值有大约 70%落在范围$[(\bar{z}-\sigma),(\bar{z}+\sigma)]$内，有大约 95%落在范围$[(\bar{z}-2\sigma),(\bar{z}+2\sigma)]$内。

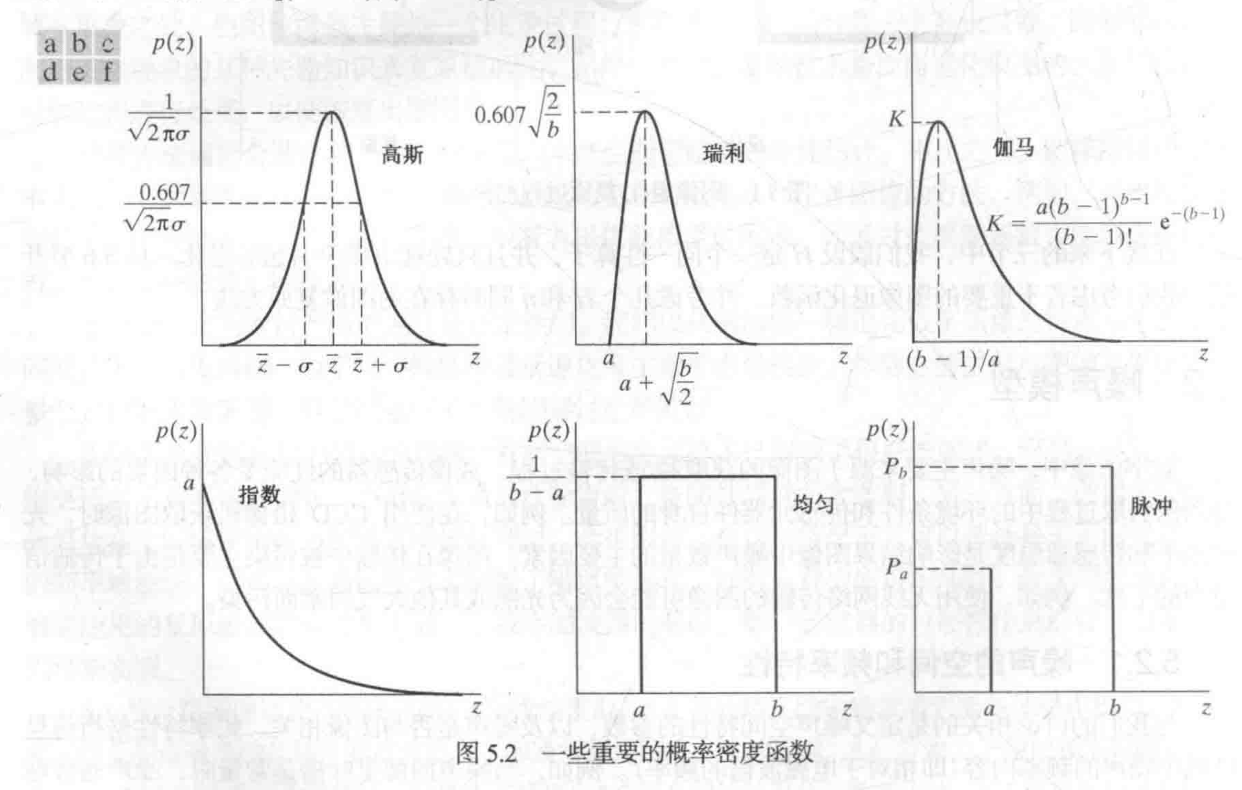


图 5.2　一些重要的概率密度函数

瑞利噪声

瑞利噪声的 PDF 由下式给出：

336 ~ 337

① 在本节中我们使用 $\bar{z}$ 而非 m 来表示均值的原因是，避免我们在后面使用 m 和 n 来表示邻域大小时引起混淆。

$$p(z)=\begin{cases}\dfrac{2}{b}(z-a)\mathrm{e}^{-(z-a)^2/b}, & z\geqslant a\\ 0, & z<a\end{cases} \tag{5.2-2}$$

概率密度的均值和方差由

$$\overline{z}=a+\sqrt{\pi b/4} \tag{5.2-3}$$

和

$$\sigma^2=\frac{b(4-\pi)}{4} \tag{5.2-4}$$

给出。图5.2(b)显示了瑞利密度的曲线。注意，距原点的位移和密度的基本形状向右变形了这一事实。瑞利密度对于近似歪斜的直方图十分适用。

爱尔兰(伽马)噪声

爱尔兰噪声的 PDF 由下式给出：

$$p(z)=\begin{cases}\dfrac{a^b z^{b-1}}{(b-1)!}\mathrm{e}^{-az}, & z\geqslant a\\ 0, & z<a\end{cases} \tag{5.2-5}$$

式中，参数 $a>0$，b 为正整数，并且“!”表示阶乘。其概率密度的均值和方差由

$$\overline{z}=\frac{b}{a} \tag{5.2-6}$$

和

$$\sigma^2=\frac{b}{a^2} \tag{5.2-7}$$

给出。图 5.2(c)显示了伽马分布密度的曲线，尽管式(5.2-5)经常称之为伽马密度，但严格地说，这只在分母为伽马函数 $\Gamma(b)$ 时才是正确的。当分母如表达式所示时，该密度称为爱尔兰密度更合适。

指数噪声

指数噪声的 PDF 由下式给出：

$$p(z)=\begin{cases}a\mathrm{e}^{-az}, & z\geq 0\\ 0, & z<0\end{cases} \tag{5.2-8}$$

式中，$a>0$。该概率密度函数的均值和方差是

$$\overline{z}=\frac{1}{a} \tag{5.2-9}$$

和

$$\sigma^2=\frac{1}{a^2} \tag{5.2-10}$$

注意，这个 PDF 是当 $b=1$ 时爱尔兰 PDF 的特殊情况。图 5.2(d)显示了该密度函数的曲线。

均匀噪声

均匀噪声的 PDF 由下式给出：

336
~
338

$$p(z)=\begin{cases}\dfrac{1}{b-a}, & a\leqslant z\leqslant b\\ 0 & \text{其他}\end{cases} \tag{5.2-11}$$

该密度函数的均值由下式给出：

$$\bar{z}=\frac{a+b}{2} \tag{5.2-12}$$

它的方差由下式给出：

$$\sigma^2=\frac{(b-a)^2}{12} \tag{5.2-13}$$

图 5.2(e)显示了均匀密度的曲线。

脉冲(椒盐)噪声

(双极)脉冲噪声的 PDF 由下式给出：

$$p(z)=\begin{cases}P_a, & z=a\\ P_b, & z=b\\ 1-P_a-P_b, & \text{其他}\end{cases} \tag{5.2-14}$$

如果 $b>a$，则灰度级 b 在图像中将显示为一个亮点；反之，灰度级 a 在图像中将显示为一个暗点。若 P_a 或 P_b 为零，则脉冲噪声称为*单极脉冲*。如果 P_a 和 P_b 两者均不可能为零，尤其是它们近似相等时，则脉冲噪声值将类似于在图像上随机分布的胡椒和盐粉微粒。由于这个原因，双极脉冲噪声也称为*椒盐噪声*。这种类型的噪声也可以使用*散粒噪声*和*尖峰噪声*来称呼。在下面的讨论中，我们将交替使用脉冲噪声和椒盐噪声这两个术语。

噪声脉冲可以为正也可以为负。标定通常是图像数字化处理的一部分。因为，与图像信号的强度相比，脉冲污染通常较大，所以在一幅图像中脉冲噪声通常被数字化为最大值(纯黑或纯白)。这样，通常假设 a 和 b 是饱和值，从某种意义上看，在数字化图像中，它们等于所允许的最大值和最小值。由于这一结果，负脉冲以一个黑点(胡椒点)出现在图像中。由于相同的原因，正脉冲以白点(盐粒点)出现在图像中。对于一幅 8 比特图像，这通常意味着 $a=0$(黑)和 $b=255$(白)。图 5.2(f)显示了脉冲噪声的概率密度函数(PDF)。

前述的一组 PDF 为在实践中建立宽带噪声污染状态的模型提供了有用的工具。例如，在一幅图像中，高斯噪声源于诸如电子电路噪声以及由低照明度和/或高温带来的传感器噪声。瑞利密度有助于在深度成像中表征噪声现象。指数密度和伽马密度在激光成像中很有用。如前段提及的那样，脉冲噪声在快速过渡的情况下产生，如在成像期间发生的错误开关操作。均匀密度可能是实践中描述得最少的，然而，均匀密度作为仿真中使用的许多随机数生成器的基础是非常有用的(见 Peebles[1993]和 Gonzalez, Woods and Eddins[2004])。

例 5.1 噪声图像及其直方图。

图 5.3 显示了一幅非常适合于阐述刚刚讨论的噪声模型的测试图案。这是一幅适合使用的图案，因为它由简单的恒定的区域组成，且其从黑到近似于白仅有 3 个灰度级增长跨度。这方便了附加在图像上的各种噪声成分特性的视觉分析。

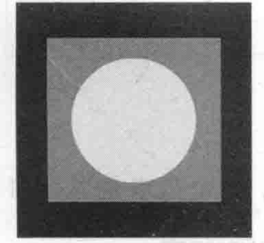


图 5.3 用于说明显示在图 5.2 中的噪声 PDF 特性的测试图案

338 ~ 339

图5.4 显示的是叠加了本节所讨论的 6 种噪声的测试图案。每幅图像下面所示的是从图像直接计算得到的直方图。在每种情况下选择了

噪声的参数，以便对应于测试图案中 3 种灰度级的直方图会开始合并。这会使得噪声十分明显，但不会遮蔽底层图像的基本结构。

比较图 5.4 中的直方图和图 5.2 中的概率密度函数，可以看到很接近的对应关系。对于椒盐噪声，示例的直方图在密度标度的白端有一个额外的尖峰，因为噪声成分为纯黑或纯白，且在测试图案中最亮的成分(圆)是亮灰度。除了少许亮度不同外，在图 5.4 中很难区分出前 5 幅图像有什么明显不同，即使它们的直方图有明显的区别。由脉冲噪声污染的图像的椒盐噪声，是唯一一种引起退化的、视觉上可区分的噪声类型。

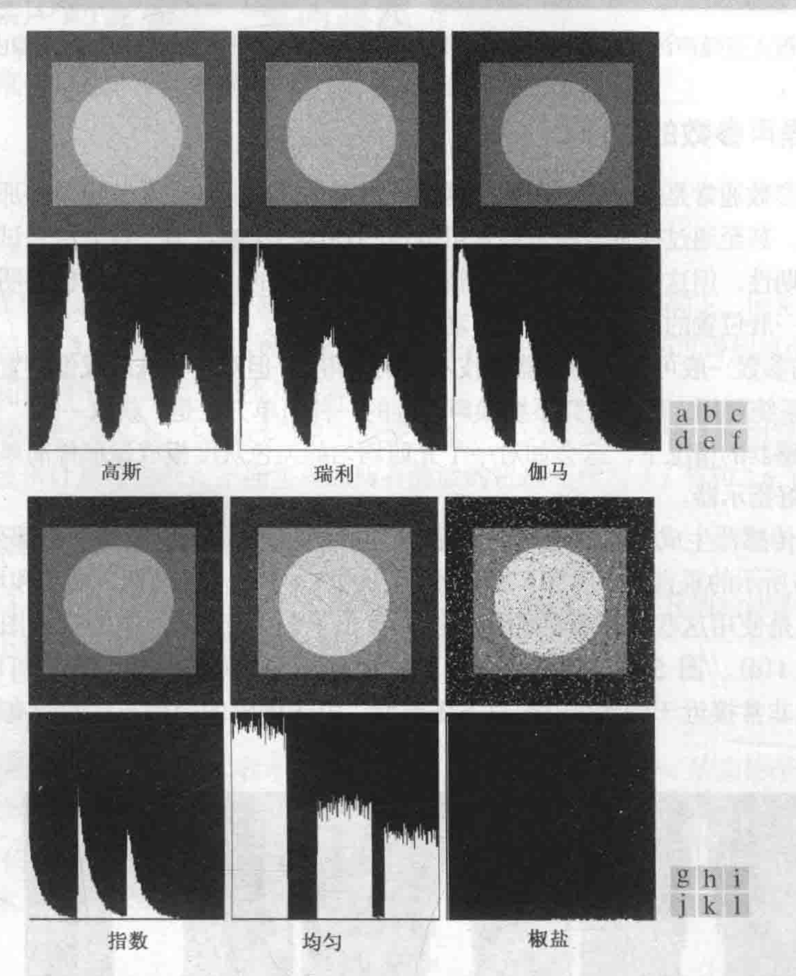


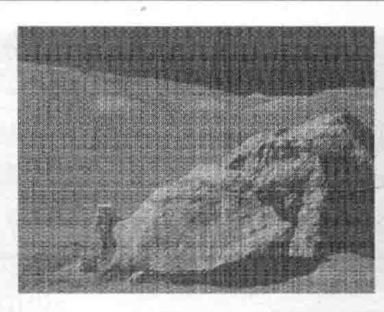
图 5.4　对图 5.3 中的图像添加高斯、瑞利、伽马、指数、均匀和椒盐噪声后的图像与直方图

5.2.3　周期噪声

图像中的周期噪声是在图像获取期间由电力或机电干扰产生的。这是在本章中讨论的唯一一种空间相关噪声。如 5.4 节所讨论的那样，周期噪声可通过频率域滤波来显著地减少。例如，考虑图5.5(a)中的图像。这幅图像被不同频率的(空间)正弦噪声严重干扰。一个纯正弦波的傅里叶变换是位于正弦波共轭频率处的一对共轭脉冲[①](见表 4.3)。因此，如果空间域中正弦波的振幅足够强，我们在该图像的谱中将看到图像中每个正弦波的脉冲对。如图5.5(b)所示，确实如此，由于在这个特殊例子中频率值是这样安排的，所以脉冲以一个近似的圆出现。在 5.4 节中我们将进一步讨论此问题，以及周期噪声的其他例子。

340
～
341

① 注意不要混淆频率域中的术语脉冲与脉冲噪声中使用的同一术语。



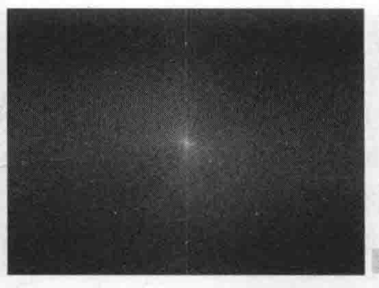


图 5.5 (a)被正弦噪声污染的图像；(b)谱(每对共轭脉冲对应于一个正弦波)(原图像由 NASA 提供)

5.2.4 噪声参数的估计

周期噪声的参数通常是通过检测图像的傅里叶谱来估计的。如前几节提及的那样，周期噪声趋向于产生频率尖峰，甚至通过视觉分析也经常可以检测到这些尖峰。另一种方法是试图直接由图像推断出噪声成分的周期性，但这仅在非常简单的情况下才是可能的。在噪声尖峰格外明显或可以使用关于干扰的频率成分一般位置的某些知识时，自动分析是可能的。

噪声 PDF 的参数一般可以从传感器的技术说明中得知，但对于特殊的成像装置通常需要估计这些参数。如果成像系统可用，那么研究系统噪声特性的一种简单方法是，获取一组“平坦”环境的图像。例如，在光学传感器的情况下，这会如对一个光照均匀的纯色灰度板成像那样简单。结果图像是典型的系统噪声的良好指示器。

当仅有通过传感器生成的图像可用时，通常可由合理的恒定灰度值的一小部分来估计 PDF 的参数。例如，图5.6所示的垂直条带(150×20 像素)是从图5.4中所示的高斯、瑞利和均匀图像中截取的。所显示的直方图是使用这些小条带的图像数据计算出来的。与图 5.6 中的直方图对应的图 5.4 中的直方图，是图 5.4(d)、图 5.4(e)和图 5.4(k)这三组直方图中间的一组。我们可以看出，这些对应的直方图的形状非常接近于图 5.6 中的直方图形状。由于缩放的原因，它们的高度不同，但形状明显类似。

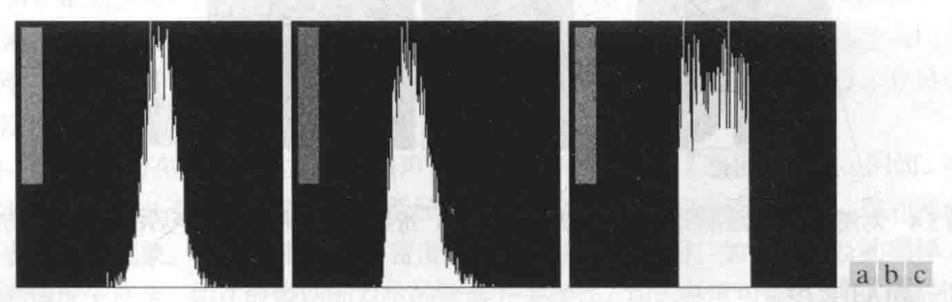


图 5.6 使用小条带(显示为插图)从图 5.4 中的(a)高斯、(b)瑞利和(c)均匀噪声图像计算的直方图

来自图像条带数据的最简单的用途是计算灰度级的均值和方差。考虑由 S 表示的一个条带(子图像)，并令 $p_S(z_i)$，$i=0, 1, 2, \cdots, L-1$ 表示 S 中像素灰度的概率估计(归一化直方图值)，其中 L 是整个图像中可能的灰度数(例如，对于 8 比特图像，L 为 256)。如第 3 章那样，我们估计 S 的均值和方差如下：

$$\overline{z} = \sum_{i=0}^{L-1} z_i p_S(z_i) \tag{5.2-15}$$

$$\sigma^2 = \sum_{i=0}^{L-1} (z_i - \overline{z})^2 p_S(z_i) \tag{5.2-16}$$

341 ~ 343

直方图的形状确定最接近的 PDF 匹配。如果其形状近似于高斯，那么均值和方差正是我们所需要的，因为高斯 PDF 完全由这两个参数指定。对于 5.2.2 节讨论的其他形状，我们使用均值和方差来解出参数 a 和 b。使用不同的方法来处理脉冲噪声，因为需要的估计是黑、白像素出现的实际概率。获得这些估计值需要黑白像素可见，因此为了计算直方图，图像中需要有一个相对恒定的中等灰度区域。对应于黑、白像素的尖峰的高度是式(5.2-14)中 P_a 和 P_b 的估计值。

5.3　只存在噪声的复原——空间滤波

当一幅图像中唯一存在的退化是噪声时，式(5.1-1)和式(5.1-2)变成

$$g(x,y)=f(x,y)+\eta(x,y) \tag{5.3-1}$$

和

$$G(u,v)=F(u,v)+N(u,v) \tag{5.3-2}$$

噪声项是未知的，故从 $g(x, y)$或 $G(u, v)$中减去它们不现实。在周期噪声的情况下，通常可由 $G(u, v)$的谱来估计 $N(u, v)$，正如 5.2.3 节所述。在这种情况下，从 $G(u, v)$中减去 $N(u, v)$可得到原图像的一个估计。然而，这种类型的知识通常只是例外而不是规律。

在仅存在加性噪声的情况下，可以选择空间滤波方法。空间滤波已在第 3 章中详细讨论过。除了执行一种特定的滤波来计算特性外，下面实现的所有滤波的机理完全与 3.5 节和 3.6 节中讨论的一样。

5.3.1　均值滤波器

本节简要讨论 3.5 节中介绍过的空间滤波器的降噪能力，并探讨一些其他的滤波器，这些滤波器的性能在许多情况下要优于 3.5 节中讨论的滤波器的性能。

算术均值滤波器

这是最简单的均值滤波器。令 S_{xy} 表示中心在点(x, y)处、大小为 $m\times n$ 的矩形子图像窗口(邻域)的一组坐标。算术均值滤波器在 S_{xy} 定义的区域中计算被污染图像 $g(x, y)$ 的平均值。在点(x, y)处复原图像 $\hat{f}$ 的值，就是简单地使用 S_{xy} 定义的区域中的像素计算出的算术均值，即

假设 m 和 n 是奇整数。

$$\hat{f}(x,y)=\frac{1}{mn}\sum_{(s,t)\in S_{xy}} g(s,t) \tag{5.3-3}$$

这个操作可以使用大小为 $m\times n$ 的一个空间滤波器来实现，滤波器的所有系数均为其值的 $1/mn$。均值滤波平滑一幅图像中的局部变化，虽然模糊了结果，但降低了噪声。

几何均值滤波器

使用几何均值滤波器复原的一幅图像由如下表达式给出：

$$\hat{f}(x,y)=\left[\prod_{(s,t)\in S_{xy}} g(s,t)\right]^{\frac{1}{mn}} \tag{5.3-4}$$

其中，每个复原的像素由子图像窗口中像素的乘积的 $1/mn$ 次幂给出。如例 5.2 所示，几何均值滤波器实现的平滑可与算术均值滤波器相比，但这种处理中丢失的图像细节更少。

343 ~ 345

谐波均值滤波器

谐波均值滤波操作由如下表达式给出：

$$\hat{f}(x, y) = \frac{mn}{\sum_{(s,t)\in S_{xy}} \frac{1}{g(s,t)}} \tag{5.3-5}$$

谐波均值滤波器对于盐粒噪声效果较好，但不适用于胡椒噪声。它善于处理像高斯噪声那样的其他噪声。

逆谐波均值滤波器

逆谐波均值滤波器基于如下表达式产生一幅复原的图像：

$$\hat{f}(x, y) = \frac{\sum_{(s,t)\in S_{xy}} g(s,t)^{Q+1}}{\sum_{(s,t)\in S_{xy}} g(s,t)^{Q}} \tag{5.3-6}$$

其中 Q 称为滤波器的阶数。这种滤波器适合减少或在实际中消除椒盐噪声的影响。当 Q 值为正时，该滤波器消除胡椒噪声；当 Q 值为负时，该滤波器消除盐粒噪声。但它不能同时消除这两种噪声。注意，当 $Q=0$ 时，逆谐波均值滤波器简化为算术均值滤波器；而当 $Q=-1$ 时，则为谐波均值滤波器。

例 5.2　均值滤波器的说明。

图5.7(a)显示了一块电路板的 8 比特 X 射线图像，图5.7(b)显示了同一幅图像，但该图像已被均值为零、方差为 400 的加性高斯噪声污染。对于这种类型的图像，这是非常严重的噪声。图5.7(c)和图5.7(d)分别显示了使用大小为 3 × 3 的算术均值滤波器和同样大小的几何均值滤波器滤除噪声后的结果。尽管这两种噪声滤波器都衰减了噪声，但几何均值滤波器并未像算术均值滤波器那样使图像变得模糊。例如，图像顶部的连接片在图5.7(d)中比在图5.7(c)中更为清晰。图像的其他部分同样如此。

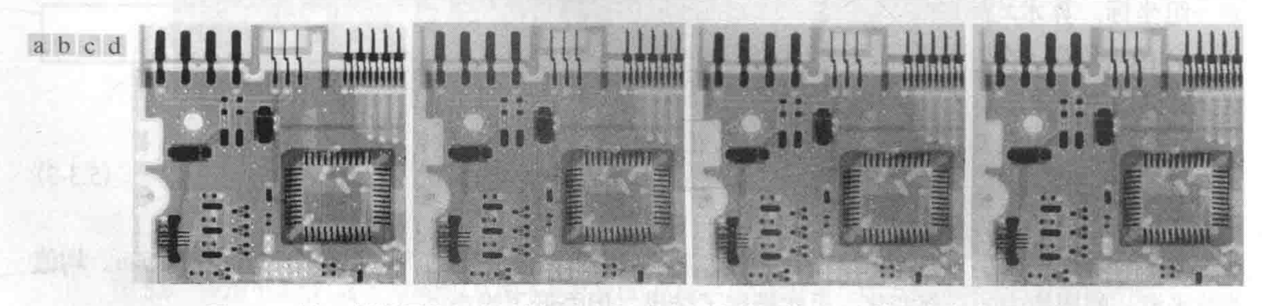


图 5.7　(a) X 射线图像；(b)被加性高斯噪声污染的图像；(c)使用大小为 3×3 的算术均值滤波器对图像滤波后的结果；(d)使用相同大小的几何均值滤波器对图像滤波后的结果(原图像由 Lixi 公司的 Joseph E. Pascente 先生提供)

图5.8(a)显示了同样的电路图像，但被概率为 0.1 的胡椒噪声污染。类似地，图5.8(b)显示了已被具有相同概率的盐粒噪声污染了的图像。图5.8(c)显示了使用 $Q=1.5$ 的逆谐波均值滤波器对图5.8(a)滤波的结果。图5.8(d)显示了使用 $Q=-1.5$ 的逆谐波均值滤波器对图5.8(b)滤波后的结果。两种滤波器都有很好的去噪效果。这种正阶滤波器除了使得暗区稍微有些淡化和模糊之外，都使背景变得更为清晰。负阶滤波器的作用正好相反。

345 ~ 346

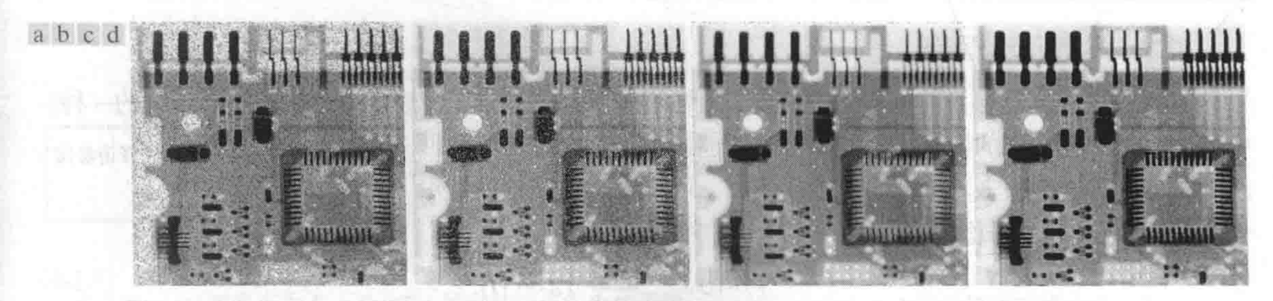


图 5.8　(a) 被概率为 0.1 的胡椒噪声污染的图像；(b) 被相同概率的盐粒噪声污染的图像；(c) 使用大小为 3×3、阶数为 1.5 的逆谐波均值滤波器对图 (a) 滤波后的结果；(d) 使用 $Q = -1.5$ 的逆谐波均值滤波器对图 (b) 滤波后的结果

总之，算术均值滤波器和几何均值滤波器（尤其是后者）更适合于处理高斯或均匀随机噪声。逆谐波均值滤波器更适合于处理脉冲噪声，但它还有一个缺点，即必须知道噪声是暗噪声还是亮噪声，以便为 Q 选择的正确符号。如果 Q 的符号选择错误，则可能会引起灾难性的后果，如图5.9 所示。下面几节讨论的一些滤波器可消除这种缺点。

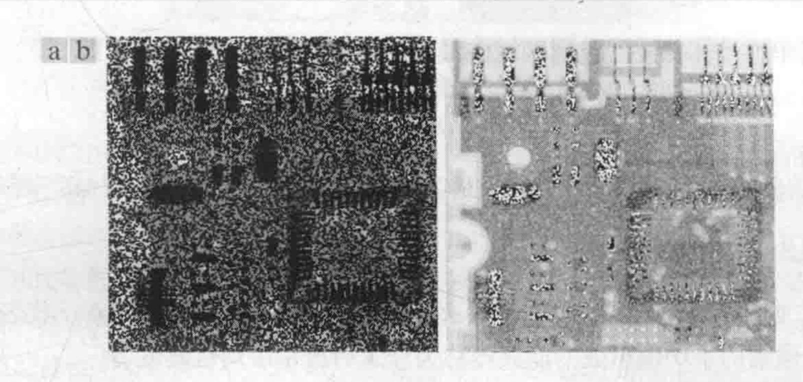


图 5.9　在逆谐波滤波中选择错误符号的结果：(a) 使用大小为 3×3 且 $Q = -1.5$ 的逆谐波滤波器对图5.8 (a) 滤波后的结果；(b) 当 $Q = 1.5$ 时对图5.8 (b) 滤波后的结果

5.3.2　统计排序滤波器

3.5.2 节介绍过统计排序滤波器。我们现在扩充那一节的讨论，并介绍其他一些统计排序滤波器。正如3.5.2 节所述，统计排序滤波器是空间域滤波器，空间域滤波器的响应基于由该滤波器包围的图像区域中的像素值的顺序（排序）。排序结果决定滤波器的响应。

中值滤波器

最著名的统计排序滤波器是中值滤波器，如其名称所示的那样，它使用一个像素邻域中的灰度级的中值来替代该像素的值，即

$$\hat{f}(x, y) = \underset{(s,t)\in S_{xy}}{\text{median}}\{g(s, t)\} \tag{5.3-7}$$

在(x, y)处的像素值是计算的中值。中值滤波器的应用非常普遍，因为对于某些类型的随机噪声，它们可提供良好的去噪能力，且与相同尺寸的线性平滑滤波器相比，引起的模糊更少。在存在单极或双极脉冲噪声的情况下，中值滤波器尤其有效。事实上，如下面的例5.3 所示，中值滤波器对于这种噪声污染的图像可得到非常好的处理效果。这种滤波器中值的计算和实现已在3.5.2 节中进行了详细讨论。

347 ~ 348

最大值和最小值滤波器

尽管中值滤波器是目前为止图像处理中最常用的一种统计排序滤波器，但它绝不是唯一的一种。中值相当于顺序排列的数组中间的那个数，但我们从基本统计学可回想到排列本身还有很多其他的可能性。例如，可以使用序列中的最后一个数值，称为最大值滤波器，由下式给出：

> 关于数的排列顺序的注释请参阅10.3.5节中的第二条注释。

$$\hat{f}(x, y) = \max_{(s,t)\in S_{xy}} \{g(s,t)\} \tag{5.3-8}$$

这种滤波器对于发现图像中的最亮点非常有用。同样，因为胡椒噪声的值非常低，作为子图像区域 S_{xy} 中这种最大值选择过程的结果，可以用这种滤波器降低它。

选择起始值的滤波器称为最小值滤波器，它由下式给出：

$$\hat{f}(x, y) = \min_{(s,t)\in S_{xy}} \{g(s,t)\} \tag{5.3-9}$$

这种滤波器对于发现图像中的最暗点非常有用。同样，作为最小值操作的结果，它可以降低盐粒噪声。

中点滤波器

中点滤波器简单地计算滤波器包围区域中最大值和最小值之间的中点，即

$$\hat{f}(x, y) = \frac{1}{2}\left[\max_{(s,t)\in S_{xy}} \{g(s,t)\} + \min_{(s,t)\in S_{xy}} \{g(s,t)\}\right] \tag{5.3-10}$$

注意，这种滤波器结合了统计排序和求平均。它最适用于处理随机分布的噪声，如高斯噪声或均匀噪声。

修正的阿尔法均值滤波器

假设在邻域 S_{xy} 内去掉 $g(s, t)$最低灰度值的 $d/2$ 和最高灰度值的 $d/2$。令 $g_r(s, t)$代表剩下的 $mn-d$ 个像素。由这些剩余像素的平均值形成的滤波器称为修正的阿尔法均值滤波器：

$$\hat{f}(x, y) = \frac{1}{mn-d}\sum_{(s,t)\in S_{xy}} g_r(s,t) \tag{5.3-11}$$

式中，d 的取值范围可为 0 到 $mn-1$。当 $d=0$ 时，修正的阿尔法均值滤波器退化为前一节讨论的算术均值滤波器。如果选择 $d=mn-1$，则修正的阿尔法均值滤波器将退化为中值滤波器。当 d 取其他值时，修正的阿尔法均值滤波器在包括多种噪声的情况下很有用，如混合有高斯噪声和椒盐噪声的情况。

例 5.3 统计排序滤波器的说明。

图5.10(a)显示了被概率为 $P_a = P_b = 0.1$ 的椒盐噪声污染的电路板图像。图5.10(b)显示了用大小为 3×3 的中值滤波器滤波的结果。对图 5.10(a)的改进显而易见，但一些噪声点仍然可见。使用中值滤波器[对图 5.10(b)中的图像]进行第二次滤波处理后，去掉了大部分这样的噪声点，仅剩下非常少的可见噪声点。这些噪声点在经过第三次中值滤波处理后全部消除。这些结果是说明中值滤波器处理脉冲型加性噪声能力的很好例子。记住，使用中值滤波器对图像重复地进行处理会使图像变模糊，所以希望保持尽可能低的处理次数。

图5.11(a)显示了把最大值滤波器用于图5.8(a)所示“胡椒”噪声图像的结果。这种滤波器对于去除图像中的“胡椒”噪声的确很合适，但我们注意到它同时也从黑色物体的边缘去除了一些黑色像素(即将这些像素设置为亮灰度级)。图 5.11(b)显示了把最小值滤波器用于图 5.8(b)的结果。在这种情况下，在噪声消除方面，最小值滤波器的确比最大值滤波器更好，但它同时也从明亮物体的边缘去除了一些白色像素。这样就使亮物体变小，而同时使暗物体变大(就像图像顶部的连接片那样)，因为围绕这些物体的白点被设置成了暗灰度级。

348 ~ 351

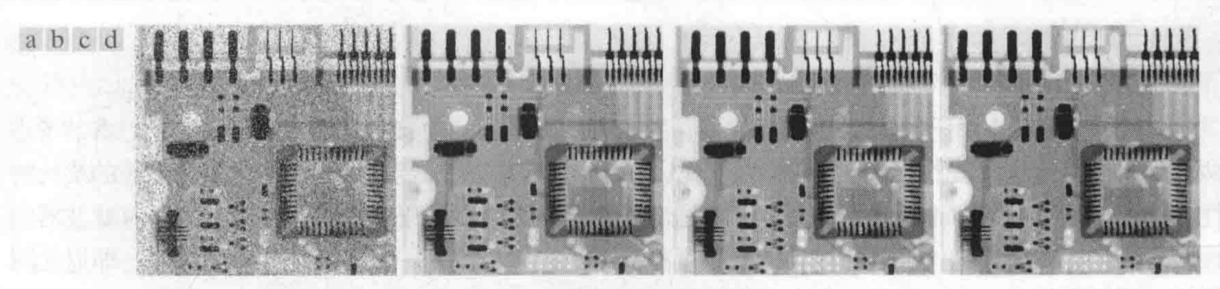


图 5.10　(a) 被概率为 $P_a = P_b = 0.1$ 的椒盐噪声污染的图像；(b) 使用大小为 3×3 的中值滤波器对图像滤波一次后的结果；(c) 使用该滤波器对图 (b) 滤波后的结果；(d) 使用相同的滤波器处理图 (c) 后的结果

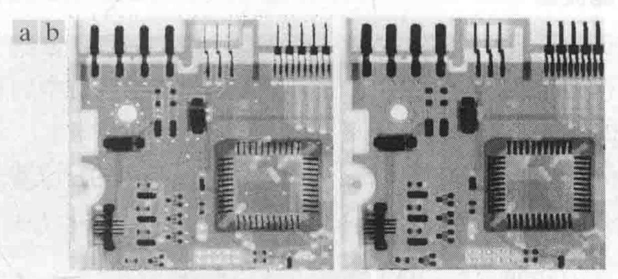


图 5.11　(a) 使用大小为 3×3 的最大值滤波器对图 5.8 (a) 滤波后的结果；(b) 使用相同大小的最小值滤波器对图 5.8 (b) 滤波后的结果

下面说明修正的阿尔法均值滤波器。图5.12 (a) 显示了一幅被均值为零、方差为 800 的加性均匀噪声污染的图像。这是一种高级别的噪声污染，它被 $P_a = P_b = 0.1$ 的椒盐噪声叠加而进一步劣化，如图5.12 (b) 所示。对这幅图像中的高级别噪声使用更大的滤波器是正确的。图5.12 (c) 到图5.12 (f) 依次显示了使用大小为 5×5 的算术均值、几何均值、中值和 $d=5$ 的修正的阿尔法均值滤波器处理后的图像。如所预料的那样，由于脉冲噪声的存在，算术均值滤波器和几何均值滤波器(尤其是后者)并未起到良好的作用。中值滤波器和修正的阿尔法均值滤波器则要好得多，在降噪方面修正的阿尔法均值滤波器的效果更好一些。例如，注意图像左上角第 4 个连接片在修正的阿尔法均值滤波器处理后变得稍微平滑一些。这并不意外，因为对于高的 d 值，修正的阿尔法均值滤波器接近于中值滤波器的性能，但仍然保留了一些平滑能力。

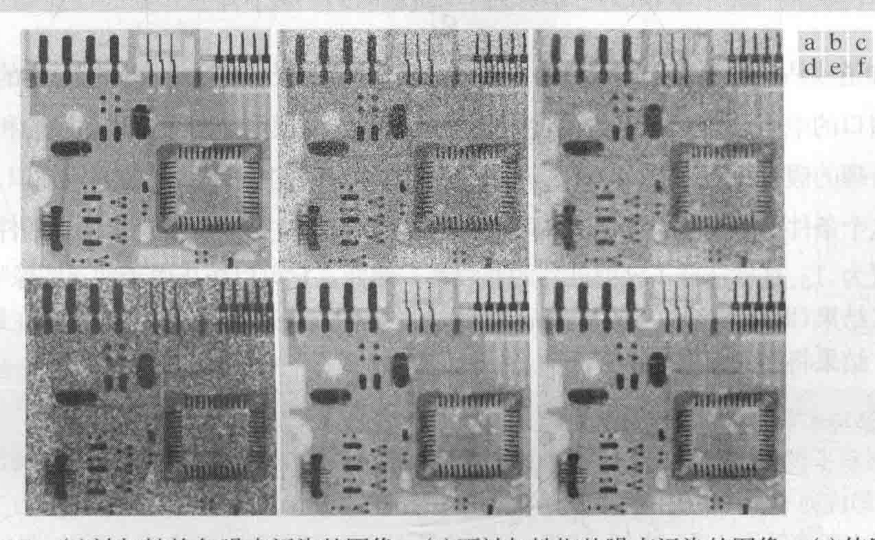


图 5.12　(a) 被加性均匀噪声污染的图像；(b) 再被加性椒盐噪声污染的图像；(c) 使用大小为 5×5 的算术均值滤波器对图 (b) 滤波后的结果；(d) 使用大小为 5×5 的几何均值滤波器对图 (b) 滤波后的结果；(e) 使用大小为 5×5 的中值滤波器对图 (b) 滤波后的结果；(f) 使用大小为 5×5 且 $d=5$ 的修正的阿尔法均值滤波器对图 (b) 滤波后的结果

350 ~ 352

5.3.3 自适应滤波器

迄今为止讨论的用于图像处理的滤波器，并未考虑图像中的一点对其他点的特征变化。本节考虑两个简单的自适应滤波器，滤波器的特性变化以 $m \times n$ 矩形窗口 S_{xy} 定义的滤波器区域内图像的统计特性为基础。如下面的讨论所表明的那样，自适应滤波器的性能要优于迄今为止讨论过的所有滤波器的性能。改善滤波能力的代价是滤波器的复杂度提高了。记住，我们一直在处理这样的情况，即退化图像等于原始图像加噪声，但还没有考虑其他类型的退化。

自适应局部降低噪声滤波器

随机变量最简单的统计度量是其均值和方差。作为自适应滤波器的基础，它们是合理的参数，因为它们是与图像外观紧密相关的量。均值给出了在其上计算均值的区域中的平均灰度的度量，而方差则给出了该区域的对比度的度量。

滤波器作用于局部区域 S_{xy}。滤波器在该区域中心任意一点(x, y)上的响应基于以下 4 个量：(a) $g(x, y)$，带噪图像在点(x, y)上的值；(b) σ_η^2，污染 $f(x, y)$ 以形成 $g(x, y)$的噪声的方差；(c) m_L，S_{xy} 中像素的局部均值；(d) σ_L^2，S_{xy} 中像素的局部方差。我们希望滤波器的性能如下：

1. 如果 σ_η^2 为零，则滤波器应该简单地返回 $g(x, y)$的值。这无关紧要，在零噪声情况下 $g(x, y)$ 等于 $f(x, y)$。
2. 如果局部方差与 σ_η^2 是高度相关的，则滤波器返回 $g(x, y)$的一个近似值。典型地，高局部方差与边缘相关，并且应该保护这些边缘。
3. 如果两个方差相等，我们则希望滤波器返回 S_{xy} 中像素的算术均值。这种情况发生在局部区域与整个图像有相同特性的条件下，并且局部噪声将通过简单地求平均来降低。

基于这些假设得到的 $\hat{f}(x, y)$ 的自适应表达式可以写成

$$\hat{f}(x, y) = g(x, y) - \frac{\sigma_\eta^2}{\sigma_L^2}\left[g(x, y) - m_L\right] \tag{5.3-12}$$

唯一需要知道或估计的量是全部噪声的方差 σ_η^2。其他参数要从每个(x, y)处的 S_{xy} 中的像素来计算，(x, y)是滤波器窗口的中心。式(5.3-12)中隐含的假设为 $\sigma_\eta^2 \leqslant \sigma_L^2$。模型中的噪声是加性的和位置无关的，因此这是一个合理的假设，因为 S_{xy} 是 $g(x, y)$的子集。然而，我们很少有 σ_η^2 的确切知识，因此在实际中很可能违反这个条件。由于这个原因，式(5.3-12)的实现应建立一个测试，以便当条件 $\sigma_\eta^2 > \sigma_L^2$ 发生时，把比率设置为 1。这就造成了该滤波器的非线性。然而，它可以防止由于缺乏图像噪声方差的知识而产生无意义结果(即负灰度级，取决于 m_L 的值)。另一种方法是允许出现负值，并在最后重新标定灰度值。因而，结果将损失图像的动态范围。

例 5.4 自适应局部降低噪声滤波的说明。

图5.13(a)显示了被均值为零、方差为 1000 的加性高斯噪声污染的电路板图像。这是噪声污染水平严重的图像，但它却成为了一个比较相应滤波器性能的理想测试基础。图5.13(b)是使用大小为 7×7 的算术均值滤波器处理噪声图像后的结果。噪声被平滑，但代价是图像被严重模糊。类似的解释也适用于图 5.13(c)，它显示了使用大小仍为 7×7 的几何均值滤波器处理噪声图像后的结果。这两幅滤波后的图像间的区别与我们在例 5.2 中讨论过的类似，仅存在模糊程度的不同。

352 ~ 353

图 5.13(d)显示了使用式(5.3-12)且 $\sigma_\eta^2 = 1000$ 的自适应滤波的结果。与前两个滤波器相比，该结果的

改进很大。从总体噪声减少情况来看，自适应滤波器所达到的效果与算术和几何均值滤波器的相似。然而，使用自适应滤波器滤波后的图像更清晰一些。例如，在图5.13(d)中，图像顶部的连接片更为清晰。其他特性，如孔和图像左下方较暗部件的 8 根引脚等，在图5.13(d)中更加清楚。这些结果是使用自适应滤波器所能达到的典型结果。如在早些时候提到过的那样，改进性能的代价是增加了滤波器的复杂性。

上面的结果中对 σ_η^2 用了一个值，该值准确地与噪声方差匹配。若该值未知，且使用了太低的估计值，则算法会因为校正量比应有的小而返回与原图像非常接近的图像。估计值太高则会造成方差的比率在 1.0 处被削平，并且算法会比正常情形下更频繁地从图像中减去平均值。如果允许为负值，且图像在最后被重新标定，则如前所述，结果将损失动态范围。

a b c d

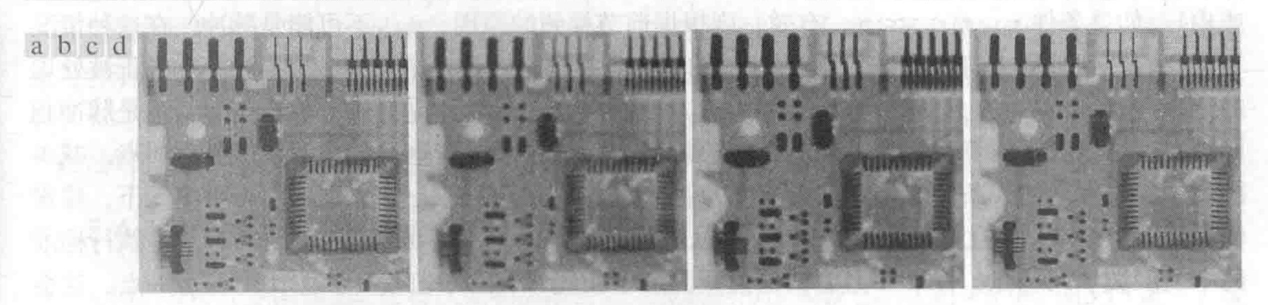


图 5.13　(a)被均值为零、方差为 1000 的加性高斯噪声污染的图像；(b)算术均值滤波的结果；(c)几何均值滤波的结果；(d)自适应降噪滤波的结果。所有滤波器的大小均为 7×7

自适应中值滤波器

对于 5.3.2 节讨论的中值滤波器，只要脉冲噪声的空间密度不大，性能就会很好(根据经验，P_a 和 P_b 小于 0.2)。本节将证明，自适应中值滤波可以处理具有更大概率的脉冲噪声。自适应中值滤波器的另一个优点是，平滑非脉冲噪声时会试图保留细节，这是传统中值滤波器所做不到的。正如前面几节中讨论的所有滤波器一样，自适应中值滤波器也工作在矩形窗口区域 S_{xy} 内。然而，与这些滤波器不同的是，自适应中值滤波器在进行滤波处理时，会根据本节列举的某些条件改变(或增大) S_{xy} 的尺寸。记住，滤波器的输出是单个数值，该值用于代替点(x, y)处的像素值，点(x, y)是给定时刻窗口 S_{xy} 的中心。

考虑如下符号：

$z_{\min} = S_{xy}$ 中的最小灰度值
$z_{\max} = S_{xy}$ 中的最大灰度值
$z_{\text{med}} = S_{xy}$ 中的灰度值的中值
$z_{xy} =$ 坐标(x, y)处的灰度值
$S_{\max} = S_{xy}$ 允许的最大尺寸

自适应中值滤波算法以两个进程工作，分别表示为进程 A 和进程 B，如下所示：

进程 A：
$A_1 = z_{\text{med}} - z_{\min}$
$A_2 = z_{\text{med}} - z_{\max}$
如果 $A_1 > 0$ 且 $A_2 < 0$，则转到进程 B
否则增大窗口尺寸
如果窗口尺寸 $\leqslant S_{\max}$，则重复进程 A
否则输出 z_{med}

353~355

进程 B：$B_1 = z_{xy} - z_{min}$

$B_2 = z_{xy} - z_{max}$

如果 $B_1 > 0$ 且 $B_2 < 0$，则输出 z_{xy}

否则输出 z_{med}

理解该算法的关键在于，要记住它有 3 个主要目的：去除椒盐(脉冲)噪声，平滑其他非脉冲噪声，并减少诸如物体边界细化或粗化等失真。值 z_{min} 和 z_{max} 在算法统计上认为是类脉冲噪声成分，即使它们在图像中并不是最低和最高的可能像素值。

利用这些观察结果，我们看到，进程 A 的目的是确定中值滤波器的输出 z_{med} 是否是一个脉冲(黑或白)。如果条件 $z_{min} < z_{med} < z_{max}$ 有效，则根据前节提到的原因，z_{med} 不可能是脉冲。在这种情况下，我们转到进程 B 进行测试，看窗口 z_{xy} 的中心点本身是否是一个脉冲(回忆可知 z_{xy} 是正被处理的点)。若条件 $B_1 > 0$ 且 $B_2 < 0$ 为真，则 $z_{min} < z_{xy} < z_{max}$，$z_{xy}$ 就不是脉冲，原因与 z_{med} 不是脉冲相同。在这种情况下，算法输出一个不变的像素值 z_{xy}。通过不改变这些“中间灰度级”的点，减少图像中的失真。如果条件 $B_1 > 0$ 且 $B_2 < 0$ 为假，则 $z_{xy} = z_{min}$ 或 $z_{xy} = z_{max}$。在任何一种情况下，像素值都是一个极端值，且算法输出中值 z_{med}，从进程 A 可知 z_{med} 不是噪声脉冲。最后一步是执行标准的中值滤波。问题是，标准中值滤波器使用图像中相应邻域的中值代替该图像中的每一点，这会引起不必要的细节损失。

继续上面的说明，假设进程 A 确实找到了一个脉冲(若失败则测试会将它转到进程 B)。然后，算法会增大窗口尺寸并重复进程 A。该循环会一直继续，直到算法找到一个非脉冲的中值(并跳转到进程 B)，或者达到了窗口的最大尺寸。如果达到了窗口的最大尺寸，则算法返回 z_{med} 值。注意，不能保证该值不是一个脉冲。噪声的概率 P_a 和/或 P_b 越小，或 S_{max} 在允许的范围内越大，过早退出条件发生的可能性就会越小。这似乎是合理的。随着脉冲密度的增大，我们会需要更大的窗口来消除尖峰噪声。

算法每输出一个值，窗口 S_{xy} 就被移到图像中的下一个位置。然后，算法重新初始化并应用到新位置的像素。如习题 3.18 指出的那样，仅使用新像素就可以反复更新中值，因而减少了计算开销。

例 5.5　自适应中值滤波的说明。

图5.14(a)是被概率为 $P_a = P_b = 0.25$ 的“椒盐”噪声污染的电路板图像，该噪声的概率为图5.10(a)中所用噪声概率的 2.5 倍。此处噪声水平非常高，以至于模糊了图像的大部分细节。作为比较的基础，图像首先使用最小中值滤波器进行滤波，消除大部分可见的脉冲噪声痕迹。做这种处理需要一个大小为 7×7 的中值滤波器，并且处理结果显示在图5.14(b)中。虽然噪声被有效地消除，但该滤波明显导致了图像细节的损失。例如，图像顶部的一些连接片出现了失真或断裂。其他的图像细节也同样失真了。

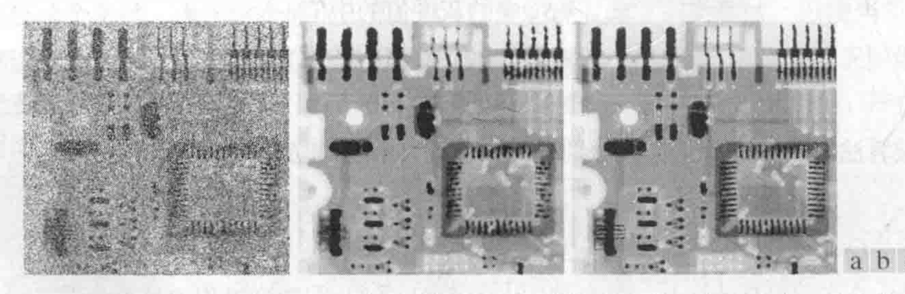


图 5.14　(a)被概率 $P_a = P_b = 0.25$ 的椒盐噪声污染的图像；(b)使用大小为 7×7 的中值滤波器的对图像滤波的结果；(c)使用 $S_{max} = 7$ 的自适应中值滤波器对图像滤波的结果

355 ~ 356

图5.14(c)是使用 $S_{max} = 7$ 的自适应中值滤波器的结果。噪声消除性能与中值滤波器相似。但是，自适应滤波器在保持清晰度和细节方面的确做得更好。连接片几乎没有失真，使用中值滤波器时被模糊或失真

而无法识别的一些特征，在图5.14(c)中变得更清晰。两个明显的例子是穿透主板的小白孔，以及图像左下四分之一处有 8 根引脚的黑色器件。

考虑到图5.14(a)中的高水平噪声，自适应算法的性能已经非常好。允许选择的最大窗口尺寸取决于应用，但合理的初始值可首先通过使用各种尺寸的标准中值滤波器进行实验来估计。这将根据自适应算法的期望性能建立直观的基准。

5.4 用频率域滤波消除周期噪声

用频率域技术可以有效地分析并滤除周期噪声。这种技术的基本概念是在傅里叶变换中，周期噪声在对应于周期干扰的频率处，以集中的能量脉冲形式出现；方法是用一个选择性滤波器分离出噪声(见 4.10 节)。在 5.4.1 节到 5.4.3 节中，为消除基本的周期噪声，将使用三种类型的选择性滤波器(4.10 节中介绍过的带阻、带通和陷波滤波器)。在 5.4.4 节中，我们还将介绍一种最佳陷波方法。

5.4.1 带阻滤波器

4.10.1 节中介绍的理想、布特沃斯和高斯带阻滤波器的传递函数总结在表 4.6 中。图5.15显示了这些滤波器的透视图，并且下面的例子说明了使用一个带阻滤波器降低周期噪声的效果。

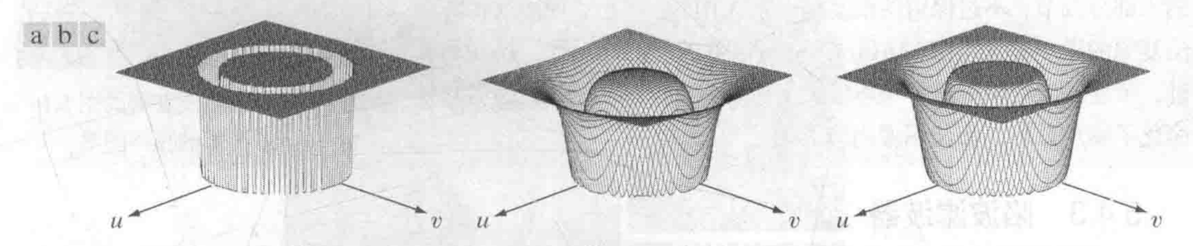


图 5.15　从左到右依次为理想滤波器、布特沃斯滤波器(阶数为 1)和高斯带阻滤波器的透视图

例 5.6　使用带阻滤波器消除周期性噪声。

带阻滤波器的主要应用之一是，在频率域噪声成分的一般位置近似已知的应用中消除噪声。一个典型的例子就是一幅被加性周期噪声污染的图像，该噪声可被近似为二维正弦函数。不难看出，一个正弦波的傅里叶变换由两个脉冲组成，它们是关于变换域坐标原点互为镜像的图像。它们的位置在表 4.3 中给出。两个脉冲实际上都是虚的(正弦曲线的傅里叶变换的实部为零)，而且彼此复共轭。我们在 5.4.3 节和 5.4.4 节中还将详述该主题。目前，我们的目的是先说明带阻滤波器。

图5.16(a)与图5.5(a)相同，显示了被不同频率的正弦噪声严重污染的图像。噪声成分可视为图5.16(b)显示的傅里叶频谱中的对称亮点对。本例中，噪声成分位于关于变换原点的近似圆上，因此使用圆对称带阻滤波器是正确的选择。图 5.16(c)显示了一个 4 阶布特沃斯带阻滤波器，它设置了适当的半径和宽度，完全包围了噪声脉冲。由于通常希望变换要尽可能小地消除细节，因此在带阻滤波中通常要求尖锐的窄滤波器。使用该滤波器对图 5.16(a)滤波的效果见图 5.16(d)，改进非常明显。即使细小的细节和纹理也被这一简单的滤波方式有效地修复了。还应注意，使用小卷积模板的直接空间域滤波方法不可能取得相同的结果。

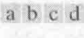
5.4.2 带通滤波器

带通滤波器执行与带阻滤波器相反的操作。4.10.1 节介绍了如何通过式(5.4-1)使用传递函数为 $H_{\mathrm{BR}}(u,v)$ 的带阻滤波器得到带通滤波器的传递函数 $H_{\mathrm{BP}}(u,v)$：

357~358

$$H_{\mathrm{BP}}(u,v)=1-H_{\mathrm{BR}}(u,v) \tag{5.4-1}$$

作为练习(见习题 5.12)，请读者根据表 4.6 中的带阻滤波器推导对应带通滤波器的表达式。



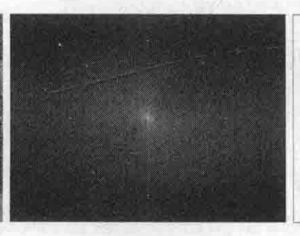
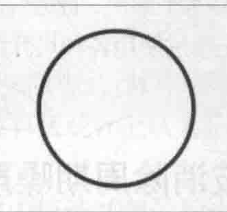





图 5.16 (a) 被正弦噪声污染的图像；(b) 图(a) 的谱；(c) 布特沃斯带阻滤波器(白色代表 1)；(d) 滤波的结果(原图像由 NASA 提供)

例 5.7 使用带通滤波提取噪声模式。

通常我们不会在一幅图像上直接执行带通滤波，因为这样会消除太多的图像细节。然而，带通滤波在一幅图像中屏蔽选中频段导致的效果时非常有用，如图5.17中的说明。产生这幅图像的方法是：(1) 应用式(5.4-1) 得到与图 5.16 中使用的带阻滤波器相对应的带通滤波器；(2) 取带通滤波变换的反变换。使用这种方法会丢失图像的大部分细节，不过保留的信息是非常有用的。显然，使用这种方法复原的噪声模式与图5.16(a) 中污染图像的噪声非常接近。换句话说，带通滤波帮助屏蔽了噪声模式。这是一个有用的结果，因为它简化了噪声分析，而与图像内容无关。

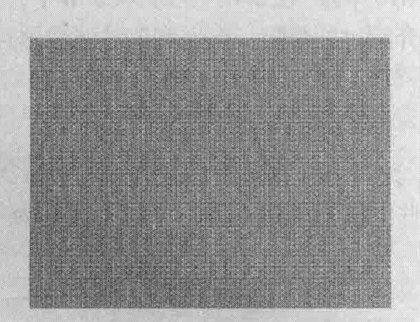


图 5.17 由带通滤波获得的图 5.16(a) 所示图像的噪声图案

5.4.3 陷波滤波器

陷波滤波器阻止(或通过)事先定义的中心频率的邻域内的频率。陷波滤波的公式详见 4.10.2 节。图5.18 分别显示了理想、布特沃斯和高斯陷波(带阻)滤波器的三维图。由于傅里叶变换的对称性，要获得有效的结果，陷波滤波器必须以关于原点对称的形式出现。这个原则的特例是，如果陷波滤波器位于原点处，在这种情况下，陷波滤波器是其本身。虽然为便于说明的目的，我们只列举了一对陷波滤波器，但可实现的陷波滤波器的对数是任意的。陷波区域的形状也可以是任意的(例如矩形)。

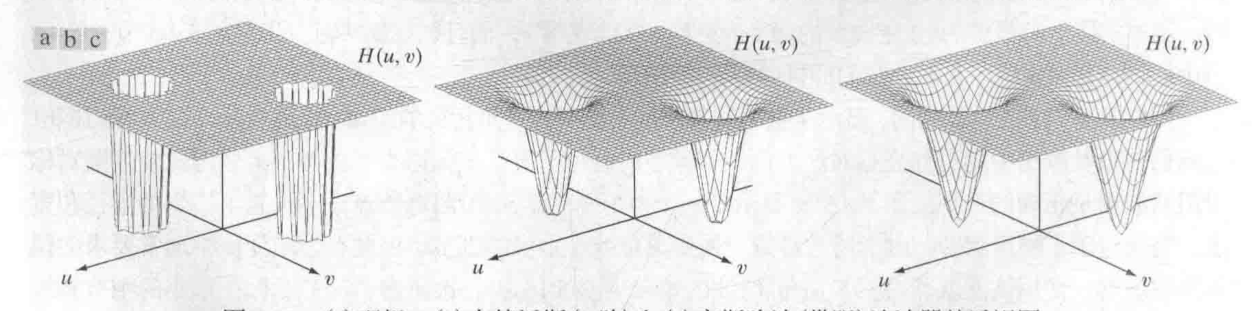


图 5.18 (a) 理想、(b) 布特沃斯(2 阶) 和(c) 高斯陷波(带阻) 滤波器的透视图

如 4.10.2 节中描述的那样，我们可以得到通过而非抑制陷波区域中所包含频率的陷波滤波器。因为这些滤波器执行与陷阻滤波器完全相反的功能，故它们的传递函数由下式给出：

$$H_{\mathrm{NP}}(u,v)=1-H_{\mathrm{NR}}(u,v) \tag{5.4-2}$$

358 ~ 359

式中，$H_{\mathrm{NP}}(u,v)$是陷波(带通) 滤波器的传递函数，这个陷波(带通) 滤波器与传递函数为$H_{\mathrm{NR}}(u,v)$的陷波(带阻) 滤波器相对应。

例 5.8　用陷波滤波器消除周期性噪声。

图 5.19(a) 显示了与图 4.51(a) 相同的图像。以下的陷波滤波方法会降低该图像中的噪声，而不会引入我们在4.8.4节中看到的明显模糊。除非该节中我们说明模糊是我们所希望的，否则找到一个合适的滤波器后，陷波滤波会更好一些。

观察图5.19(a) 中近似水平线的噪声模式，我们希望它在频率域内的成分沿垂直轴集中。然而，图 5.19(b) 所示的谱表明，噪声没有明显到具有足够清楚的模式。通过沿傅里叶变换的垂直轴构建一个简单的理想陷波带通滤波器，我们可以得到噪声成分的外表特征，如图 5.19(c) 所示。噪声模式的空间表示(陷波带通滤波结果的反变换) 见图 5.19(d)。这种噪声模式与图5.19(a) 中的模式相对应。创建一个以合理程度屏蔽噪声的合适陷波带通滤波器后，我们可由式(5.4-2) 获得相应的陷波带阻滤波器。使用陷波带阻滤波器处理图像后的结果见 5.19(e)。该图像与图5.19(a) 相比，包含有几乎看不见的噪声扫描线。

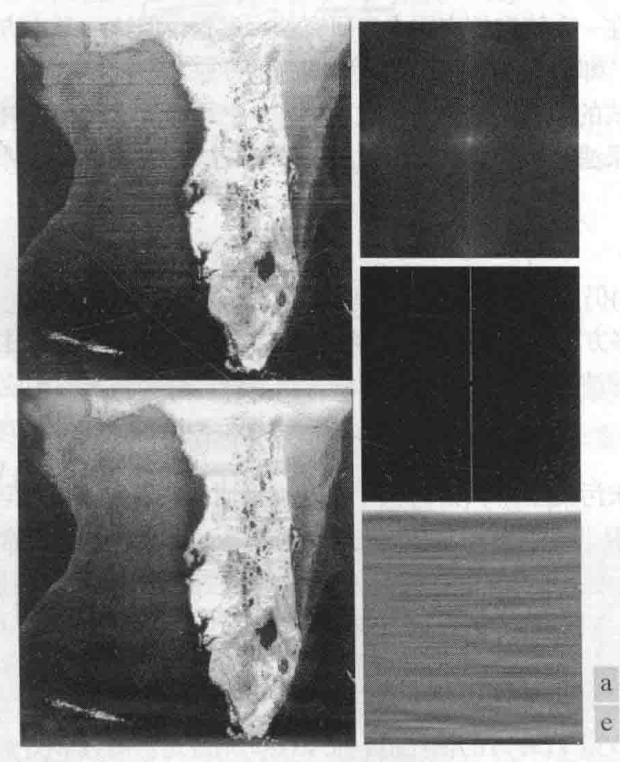


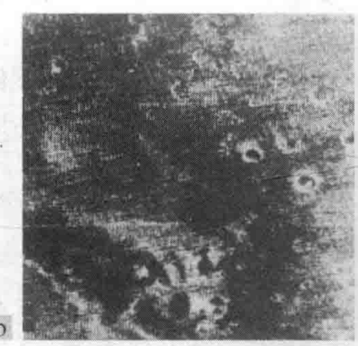
图 5.19　(a) 佛罗里达和墨西哥湾的卫星图像，显示有水平扫描线；(b) 图 (a) 的谱；(c) 叠加在图 (b) 上的陷波带通滤波器；(d) 空间噪声模式；(e) 陷波带阻滤波的结果 (原图像由 NOAA 提供)

5.4.4　最佳陷波滤波

作为周期性图像退化的另一个例子，图5.20(a) 显示了"水手 6 号"飞船拍摄的火星地形数字图像。其干扰模式与图5.16(a) 中显示的类似，但前者的模式更为精细，因而在频率平面上更难检测。图5.20(b) 显示了该问题图像的傅里叶谱。类似星形的成分是由干扰引起的，而且存在几对成分，指出该模式包含不止一个正弦成分。

存在几种干扰成分时，就不能采用前面几节讨论的方法，因为它们在滤波过程中可能会消除太多的图像信息(当图像很特别或很难获取时，很不希望出现这种现象)。另外，干扰成分通常不是单频脉冲。相反，它们通常具有携带干扰模式信息的宽边缘。从正常的变换背景中有时不容易检测到这些边缘。在许多应用中，降低这些退化影响时选择滤波方法将非常有用。这里讨论的方法是最佳的，因为在一定意义上，它最小化了复原的估计值 $\hat{f}(x,y)$ 的局部方差。

360 ~ 361




a b

图 5.20 (a)“水手 6 号”飞船拍摄的火星地形图像；(b)显示有周期干扰的傅里叶谱(图像由 NASA 提供)

该过程由两步组成，第一步是屏蔽干扰的主要成分，第二步从被污染图像中减去该模式的一个可变加权部分。虽然我们是在一个特定的应用中来研究这一过程，但这一基本方法仍然十分通用，而且能应用到其他复原工作中，即多周期性干扰是主要问题的复原工作。

第一步是提取干扰模式的主频率成分。像之前那样，这可通过在每个尖峰处放置一个陷波带通滤波器 $H_{NP}(u, v)$来完成。如果滤波器构建为只可通过与干扰模式相关的成分，那么干扰噪声模式的傅里叶变换由下式给出：

$$N(u,v) = H_{NP}(u,v)G(u,v) \tag{5.4-3}$$

其中，如通常那样，$G(u, v)$仍为被污染图像的傅里叶变换。

$H_{NP}(u, v)$的形式需要多方面判断哪些是尖峰噪声干扰。为此，通常要通过观察显示的 $G(u, v)$的频谱来交互地创建陷波带通滤波器。选择一个特殊滤波器之后，空间域中的相应模式可由下式获得：

$$\eta(x,y) = \Im^{-1}\{H_{NP}(u,v)G(u,v)\} \tag{5.4-4}$$

因为被污染图像假设是由未污染图像 $f(x,y)$ 与干扰相加形成的，若$\eta(x, y)$完全已知，则从 $g(x, y)$减去该干扰模式得到 $f(x,y)$ 将是非常简单的事情。当然，问题是这个滤波过程通常只会得到真实干扰模式的近似值。在$\eta(x,y)$的估计中，不存在的那些成分的影响可以被最小化，方法是从 $g(x,y)$中减去$\eta(x,y)$的一个加权部分来得到 $f(x,y)$ 的估计值，如下所示：

$$\hat{f}(x,y) = g(x,y) - w(x,y)\eta(x,y) \tag{5.4-5}$$

其中，如先前那样，$\hat{f}(x,y)$ 是 $f(x,y)$ 的估计值，而 $w(x,y)$则待定。函数 $w(x,y)$称为加权函数或调制函数，这个过程的目的就是选取该函数，以便以某种有意义的方法来优化结果。一种方法是选取 $w(x,y)$，使估计值 $\hat{f}(x,y)$ 在每一点(x,y)的指定邻域上的方差最小。

考虑关于点(x, y)的大小为$(2a + 1)(2b + 1)$的邻域。在坐标(x, y)处，$\hat{f}(x,y)$ 的局部方差可根据样本估计，如下所示：

$$\sigma^2(x,y) = \frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\left[\hat{f}(x+s,y+t) - \overline{\hat{f}}(x,y)\right]^2 \tag{5.4-6}$$

式中，$\overline{\hat{f}}(x,y)$ 是该邻域内 $\hat{f}$ 的平均值，即

$$\overline{\hat{f}}(x,y) = \frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\hat{f}(x+s,y+t) \tag{5.4-7}$$

362 ~ 363

在图像边缘上或接近图像边缘的点上，可用局部邻域或 0 填充的方法处理。

把式(5.4-5)代入式(5.4-6)，得

$$\sigma^2(x,y)=\frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\{[g(x+s,y+t)-w(x+s,y+t)\eta(x+s,y+t)]-[\overline{g}(x,y)-\overline{w(x,y)\eta(x,y)}]\}^2 \tag{5.4-8}$$

假设 $w(x, y)$在整个邻域内基本保持不变，则当$-a \leqslant s \leqslant a$ 和$-b \leqslant t \leqslant b$ 时，可给出近似式

$$w(x+s,y+t)=w(x,y) \tag{5.4-9}$$

这一假设在该邻域内也可得出如下表达式：

$$\overline{w(x,y)\eta(x,y)}=w(x,y)\overline{\eta}(x,y) \tag{5.4-10}$$

由这些近似，式(5.4-8)变为

$$\sigma^2(x,y)=\frac{1}{(2a+1)(2b+1)}\sum_{s=-a}^{a}\sum_{t=-b}^{b}\{[g(x+s,y+t)-w(x,y)\eta(x+s,y+t)]-[\overline{g}(x,y)-w(x,y)\overline{\eta}(x,y)]\}^2 \tag{5.4-11}$$

为最小化$\sigma^2(x,y)$，解下式得到 $w(x, y)$：

$$\frac{\partial\sigma^2(x,y)}{\partial w(x,y)}=0 \tag{5.4-12}$$

结果为

$$w(x,y)=\frac{\overline{g(x,y)\eta(x,y)}-\overline{g}(x,y)\overline{\eta}(x,y)}{\overline{\eta^2}(x,y)-\overline{\eta}^2(x,y)} \tag{5.4-13}$$

要获得复原图像 $\hat{f}(x,y)$，可根据式(5.4-13)计算 $w(x, y)$，然后使用式(5.4-5)。若 $w(x, y)$在某一邻域内假设为常量，则不必对图像中的每个 x 和 y 值计算该函数，而是在每个非重叠邻域的一点(一般为中心点)计算 $w(x, y)$，然后用它来处理该邻域内包含的所有图像点。

例 5.9　最佳陷波滤波的说明。

图5.21 到图5.23 显示了对图5.20(a)应用前述技术处理后的效果。该图像的大小为 512 × 512 像素，并且选择了 $a = b = 15$ 的邻域。图 5.21 显示了被污染图像的傅里叶谱。在这种特殊情况下，原点未移到频率平面的中心，因此 $u = v = 0$ 在图 5.21 所示变换图像的左上角。图 5.22(a)显示了 $N(u, v)$的谱，其中只存在尖峰噪声。图5.22(b)显示了取 $N(u, v)$的傅里叶反变换得到的干扰模式$\eta(x, y)$。注意，这个干扰模式与图5.20(a)中存在的噪声结构很相似。最后，图5.23 显示了应用式(5.4-5)处理后得到的图像。出于实用的目的，周期干扰已被消除。

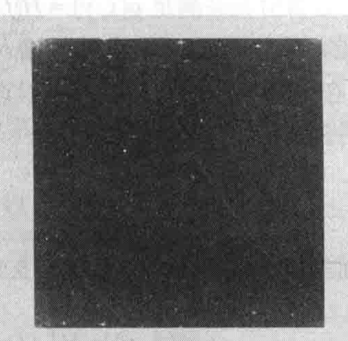


图 5.21　图 5.20(a)所示图像的(未移位的)傅里叶谱(图像由 NASA 提供)

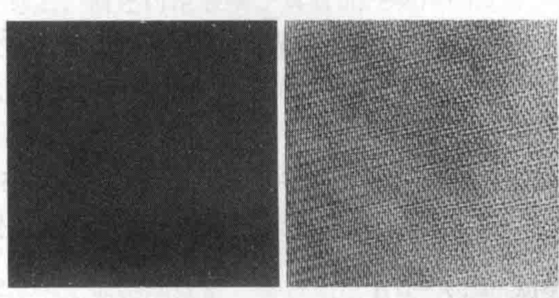


图 5.22　(a) $N(u, v)$的傅里叶谱；(b)相应的噪声干扰模式 $\eta(x, y)$(图像由 NASA 提供)

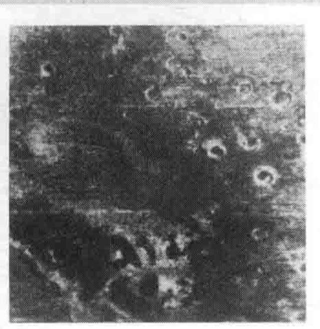


图 5.23　处理后的图像(图像由 NASA 提供)

363 ~ 365

5.5 线性、位置不变的退化

图像复原前，图 5.1 中的输入/输出关系可以表示为

$$g(x, y) = H[f(x, y)] + \eta(x, y) \tag{5.5-1}$$

现在，让我们假设 $\eta(x, y) = 0$，则 $g(x, y) = H[f(x, y)]$。基于 2.6.2 节的讨论，如果

$$H[af_1(x, y) + bf_2(x, y)] = aH[f_1(x, y)] + bH[f_2(x, y)] \tag{5.5-2}$$

则系统 H 是一个线性系统，其中 a 和 b 是标量，$f_1(x, y)$和$f_2(x, y)$是两幅输入图像。

关于线性系统理论的简要回顾，请参阅本书的网站。

若 $a = b = 1$，则式(5.5-2)变为

$$H[f_1(x,y) + f_2(x,y)] = H[f_1(x,y)] + H[f_2(x,y)] \tag{5.5-3}$$

这就是所谓的加性。这一特性简单地表明，如果 H 为线性算子，那么两个输入之和的响应等于两个响应之和。

如果 $f_2(x, y) = 0$, 则式(5.5-2)变为

$$H[af_1(x, y)] = aH[f_1(x, y)] \tag{5.5-4}$$

我们称之为均匀性。它表明任何与常数相乘的输入的响应，等于该输入响应乘以相同的常数，即一个线性算子具有加性和均匀性。

一个算子对于任意 $f(x, y)$，α 和 β，如果

$$H[f(x-\alpha, y-\beta)] = g(x-\alpha, y-\beta) \tag{5.5-5}$$

则输入/输出关系满足 $g(x, y) = H[f(x, y)]$的算子称为位置(或空间)不变系统。该定义说明图像中任意一点处的响应只取决于该点处的输入值，而与该点的位置无关。

关于连续与离散变量的讨论，请参阅 5.11.3 节的脚注。

在式(4.5-3)中，稍微(等效)改变一下离散冲激函数定义的符号，就可将 $f(x, y)$ 表示为

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha,\beta)\delta(x-\alpha, y-\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{5.5-6}$$

再次假设$\eta(x, y) = 0$，然后把式(5.5-6)代入式(5.5-1)可得

$$g(x, y) = H[f(x, y)] = H\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha, \beta)\delta(x-\alpha, y-\beta)\,\mathrm{d}\alpha\,\mathrm{d}\beta\right] \tag{5.5-7}$$

如果 H 是线性算子，且我们把加性性质扩展到积分，则有

$$g(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H[f(\alpha, \beta)\delta(x-\alpha, y-\beta)]\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{5.5-8}$$

由于 $f(\alpha, \beta)$ 与 x 和 y 无关，使用均匀性可得

$$g(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha, \beta)H[\delta(x-\alpha, y-\beta)]\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{5.5-9}$$

公式

365 ~ 366

$$h(x, \alpha, y, \beta) = H[\delta(x-\alpha, y-\beta)] \tag{5.5-10}$$

称为系统 H 的冲激响应。也就是说，在式(5.5-1)中，若$\eta(x, y) = 0$，则 $h(x, \alpha, y, \beta)$是系统 H 对坐

标(x, y)处的一个冲激的响应。在光学领域，冲激为一个光点，所以 $h(x,\alpha, y,\beta)$通常称为点扩散函数(PSF)。这一名称来自于这样一个事实，即所有物理光学系统在一定程度上会模糊(扩散)光点，模糊程度由光学部件的质量决定。

将式(5.5-10)代入式(5.5-9)，可得

$$g(x, y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha, \beta)h(x, \alpha, y, \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta \tag{5.5-11}$$

这称为第一类叠加(弗雷德霍姆)积分。该表达式是线性系统理论核心的一个基本结果。它表明，如果系统 H 对一个冲激的响应已知，则对于任意输入 $f(\alpha,\beta)$ 的响应可用式(5.5-11)来计算。换句话说，线性系统 H 完全可由其冲激响应来表征。

如果 H 是位置不变的，由式(5.5-5)可得

$$H\left[\delta(x-\alpha, y-\beta)\right]=h(x-\alpha, y-\beta) \tag{5.5-12}$$

在这种情况下，式(5.5-11)简化为

$$g(x, y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha, \beta)h(x-\alpha, y-\beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta \tag{5.5-13}$$

该式为式(4.2-20)对一个变量的卷积积分，并在习题 4.11 中推广到了二维情形。该积分告诉我们，对于任意输入 f，若已知线性系统的冲激响应，则可计算出它的响应 g。结果是冲激响应和输入函数的简单卷积。

存在加性噪声时，线性退化模型［见式(5.5-11)］的表达式变为

$$g(x, y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha, \beta)h(x, \alpha, y, \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta+\eta(x, y) \tag{5.5-14}$$

如果 H 是位置不变的，则式(5.5-14)变为

$$g(x, y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\alpha, \beta)h(x-\alpha, y-\beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta+\eta(x, y) \tag{5.5-15}$$

噪声项$\eta(x, y)$的值是随机的，并假设这些噪声项与位置无关。使用常用的卷积符号，可将式(5.5-15)写为

$$g(x,y)=h(x,y)\star f(x,y)+\eta(x,y) \tag{5.5-16}$$

或者基于卷积定理(见 4.6.6 节)，可在频率域中将它表示为

$$G(u,v)=H(u,v)F(u,v)+N(u,v) \tag{5.5-17}$$

这两个表达式与式(5.1-1)和式(5.1-2)是一致的。记住，对于离散值来说，所有的乘积是逐项相乘的。例如，$H(u, v)F(u, v)$的 ij 项是 $H(u, v)$的 ij 项和 $F(u, v)$的 ij 项的乘积。

总之，前述讨论表明，具有加性噪声的线性空间不变退化系统，可在空间域建模为退化(点扩散)函数与一幅图像的卷积，然后再加上噪声。基于卷积定理，在频率域中，同样的过程可表示为图像和退化函数的变换的乘积，然后再加上噪声的变换。在频率域中工作时，可以使用 4.11 节中讨论的一种 FFT 算法。还应记住，如 4.6.6 节所述，在实现离散傅里叶变换时，需要对函数进行填充。

许多类型的退化可近似为线性、位置不变的过程。这种方法的优点是，可以使用许多线性系统理论的工具来解决图像复原问题。与位置有关的非线性技术虽然更普遍(通常会更精确)，但它们会带来没有已知解的问题，或解决计算问题时非常困难。本章主要介绍线性、空间不变的复原技术。由于退化被建模为卷积的结果，并且图像复原试图找到应用相反过程的滤波器，所以术语图像*去卷积*通常用于表示线性图像复原。同样，用于复原处理的滤波器通常称为*去卷积滤波器*。

367 ~ 368

5.6 估计退化函数

在图像复原时，主要有3种用于估计退化函数的方法：(1)观察法，(2)试验法，(3)数学建模法。以下几节将讨论这些方法。使用以某种方式估计的退化函数来复原一幅图像的过程，有时称为盲去卷积，因为真正的退化函数很少能完全知晓。

5.6.1 图像观察估计

假设我们有一幅退化图像，而没有关于退化函数H的任何知识。基于图像被线性、位置不变的过程退化的假设，估计H的一种方法就是从图像本身来收集信息。例如，如果图像已被模糊，则我们可以观察图像中包含样本结构的一个小矩形区域，如某一物体和背景的一部分。为了降低噪声的影响，我们可以寻找一个有很强信号内容的区域(如高对比度区域)。下一步是处理子图像以便得到尽可能不模糊的结果。例如，我们可以使用锐化滤波器对这个子图像进行锐化处理，甚至可以使用手工方法处理小区域。

令$g_s(x, y)$表示要观察的子图像，令$\hat{f}_s(x, y)$表示处理过的子图像(现实中，该图像是原图像在该区域的估计图像)。然后，假设噪声的影响由于选择了一个强信号区域而可以忽略，根据式(5.5-17)可得

$$H_s(u,v)=\frac{G_s(u,v)}{\hat{F}_s(u,v)} \tag{5.6-1}$$

根据这一函数特性，然后，我们可基于位置不变的假设还原完整的退化函数$H(u, v)$。例如，假设$H_s(u, v)$的径向曲线具有高斯曲线的近似形状。我们可以利用这一信息在更大比例上构建一个有基本相同形状的函数$H(u, v)$。然后，在下面几节讨论的一种复原方法中使用$H(u, v)$。很明显，这是仅在特殊环境下使用的烦琐处理，例如复原一幅有历史价值的老照片。

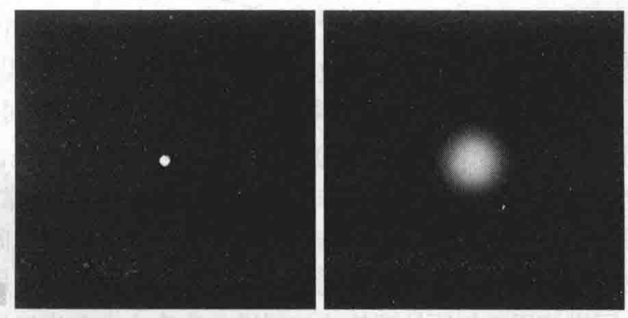
5.6.2 试验估计

如果可以使用与获取退化图像的设备相似的装置，从理论上讲，得到准确的退化估计是可能的。与退化图像类似的图像可以通过各种系统设置得到，直到这些图像退化到尽可能接近我们希望复原的程度。之后，使用相同的系统对一个冲激(小亮点)成像，得到退化的冲激响应。正如5.5节所表明的那样，线性空间不变系统完全由其冲激响应来表征。

一个冲激可由一个亮点来模拟，该点应尽可能亮，以便将噪声的影响降低到可以忽略的程度。回顾可知，由于冲激的傅里叶变换是一个常量，那么由式(5.5-17)可得

$$H(u,v)=\frac{G(u,v)}{A} \tag{5.6-2}$$

其中，与之前一样，函数$G(u, v)$是观察图像的傅里叶变换，A是一个描述冲激强度的常量。图5.24显示了一个例子。



a b

图5.24 冲激特性的退化估计：(a)一个亮冲激(放大显示)；(b)图像处理(退化)后的冲激

368 ~ 369

5.6.3 建模估计

由于退化建模能解决图像复原问题，因此多年来一直被人们使用。在某些情况下，模型甚至可以考虑引起退化的环境条件。例如，Hufnagel and Stanley[1964]根据大气湍流的物理特性提出了一个退化模型，其通式为

$$H(u,v)=\mathrm{e}^{-k(u^2+v^2)^{5/6}} \tag{5.6-3}$$

式中，k 是与湍流性质有关的常数。除了指数为 5/6 之外，该式与4.8.3 节所述高斯低通滤波的形式相同。事实上，高斯 LPF 有时用于模型淡化，以便均匀模糊。图5.25 显示了使用式(5.6-3)并取 $k=0.0025$（剧烈湍流）、$k=0.001$（中等湍流）和 $k=0.000\,25$（轻微湍流）来模拟模糊一幅图像时得到的例子。所有图像的大小均为 480×480 像素。

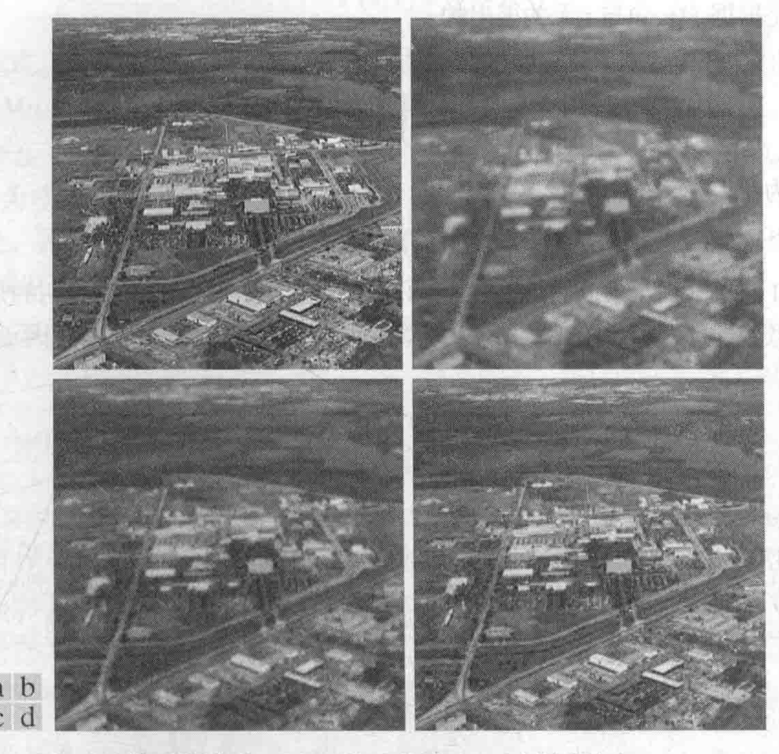


图 5.25 大气湍流模型的说明：(a) 可忽略的湍流；(b)剧烈湍流，$k=0.0025$；(c) 中等湍流，$k=0.001$；(d)轻微湍流，$k=0.000\,25$（原图像由 NASA 提供）

建模的另一种主要方法是，从基本原理开始推导一个数学模型。下面通过详细讨论一个实例来解释这一过程。在这个实例中，图像获取时被图像与传感器之间的均匀线性运动模糊了。假设图像 $f(x,y)$ 进行平面运动，$x_0(t)$和 $y_0(t)$分别是在 x 和 y 方向上随时间变化的分量。那么，记录介质（如胶片或数字存储器）上任意点的曝光总数是通过对时间间隔内瞬时曝光量的积分得到的，图像系统的快门在该时段是开着的。假设快门开关所用的时间非常短，因此光学成像过程不会受到图像运动的干扰。若设 T 为曝光时间，则有

$$g(x,y)=\int_0^T f\left[x-x_0(t),\,y-y_0(t)\right]\mathrm{d}t \tag{5.6-4}$$

式中，$g(x,y)$为模糊后的图像。

由式(4.5-7)，式(5.6-4)的傅里叶变换为

369 ~ 371

$$G(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_0^T f\left[x-x_0(t), y-y_0(t)\right]\mathrm{d}t\right]\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\,\mathrm{d}y \tag{5.6-5}$$

改变积分顺序，式(5.6-5)可写为

$$G(u,v) = \int_0^T\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\left[x-x_0(t), y-y_0(t)\right]\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\mathrm{d}x\,\mathrm{d}y\right]\mathrm{d}t \tag{5.6-6}$$

外层方括号内的积分项是位移函数$f[x-x_0(t), y-y_0(t)]$的傅里叶变换。利用式(4.6-4)可得

$$G(u,v) = \int_0^T F(u,v)\mathrm{e}^{-\mathrm{j}2\pi[ux_0(t)+vy_0(t)]}\mathrm{d}t = F(u,v)\int_0^T \mathrm{e}^{-\mathrm{j}2\pi[ux_0(t)+vy_0(t)]}\mathrm{d}t \tag{5.6-7}$$

这里，最后一步是根据$F(u,v)$与t无关得出的。

令

$$H(u,v) = \int_0^T \mathrm{e}^{-\mathrm{j}2\pi[ux_0(t)+vy_0(t)]}\,\mathrm{d}t \tag{5.6-8}$$

式(5.6-7)可表示为我们熟悉的如下形式：

$$G(u,v) = H(u,v)F(u,v) \tag{5.6-9}$$

若运动变量$x_0(t)$和$y_0(t)$已知，则可直接由式(5.6-8)得到传递函数$H(u,v)$。要说明的是，假设问题中的图像只在x方向以给定速度$x_0(t) = at/T$做匀速直线运动。当$t = T$时，图像位移的总距离为a。令$y_0(t) = 0$，由式(5.6-8)可得

$$H(u,v) = \int_0^T \mathrm{e}^{-\mathrm{j}2\pi ux_0(t)}\,\mathrm{d}t = \int_0^T \mathrm{e}^{-\mathrm{j}2\pi uat/T}\,\mathrm{d}t = \frac{T}{\pi ua}\sin(\pi ua)\mathrm{e}^{-\mathrm{j}\pi ua} \tag{5.6-10}$$

上式表明，由$u = n/a$给出的u值处，H变为零，其中n为整数。若允许y分量也随着变化，按$y_0 = bt/T$给出的运动，则退化函数变为

$$H(u,v) = \frac{T}{\pi(ua+vb)}\sin\left[\pi(ua+vb)\right]\mathrm{e}^{-\mathrm{j}\pi(ua+vb)} \tag{5.6-11}$$

如表4.3结尾处解释的那样，我们使用u和v来对式(5.6-11)取样，以便生成一个离散滤波器。

例5.10 运动引起的图像模糊。

图5.26(b)是由计算图5.26(a)中图像的傅里叶变换模糊的一幅图像，其过程是用式(5.6-11)中的$H(u,v)$乘以变换，再做反变换。图像大小为688 × 688像素，在式(5.6-11)中使用的参数为$a = b = 0.1$和$T = 1$。如5.8节和5.9节讨论的那样，从模糊图像中复原出相应的原图像会存在某些有趣的挑战，特别是当退化图像中存在噪声时。



图5.26 (a)原图像；(b)使用式(5.6-11)中的函数，且$a = b = 0.1$和$T = 1$模糊图像后的结果

371 ~ 372

5.7 逆滤波

本节介绍复原由退化函数 H 退化的图像的最初手段。退化函数已给出，或由前一节讨论的那些方法获得。最简单的复原方法是直接做逆滤波，这里我们用退化函数除退化图像的傅里叶变换 $G(u, v)$来计算原始图像傅里叶变换的估计 $\hat{F}(u,v)$，即

$$\hat{F}(u,v)=\frac{G(u,v)}{H(u,v)} \tag{5.7-1}$$

该除法是阵列操作，如 2.6.1 节中定义的那样，并联系式(5.5-17)。用式(5.1-2)的右侧替代式(5.7-1)中的 $G(u, v)$，得到

$$\hat{F}(u,v)=F(u,v)+\frac{N(u,v)}{H(u,v)} \tag{5.7-2}$$

这是一个有趣的表达式。它告诉我们，即使知道退化函数，也不能准确地复原未退化的图像［$F(u, v)$的傅里叶反变换］，因为$N(u, v)$未知。还有更糟的情况，如果退化函数是零或是非常小的值，则 $N(u, v)/H(u, v)$很容易支配估计值 $\hat{F}(u,v)$。事实上，这是经常出现的情况，下面很快会证明这一点。

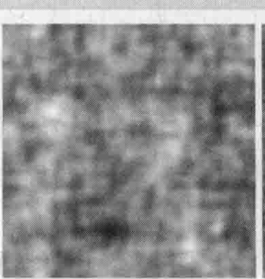
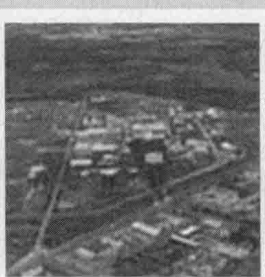
解决退化函数为零或为非常小的值的问题的一种方法是，限制滤波的频率，使其接近原点。从式(4.6-21)的讨论，我们知道 $H(0, 0)$在频率域中通常是 $H(u, v)$的最高值。因此，通过将频率限制在原点附近分析，就减少了遇到零值的概率。下面的例子中说明了这种方法。

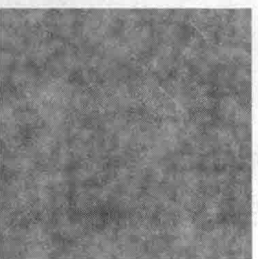
例 5.11　逆滤波。

图 5.25(b)是对产生退化图像的退化函数精确取反，并使用式(5.7-1)反向滤波的图像。也就是，使用的退化函数是

$$H(u,v)=\mathrm{e}^{-k\left[(u-M/2)^2+(v-N/2)^2\right]^{5/6}}$$

式中 $k=0.002\,5$。常数 $M/2$ 和 $N/2$ 是偏移值；它们使函数居中，以便与居中的傅里叶变换对应，就像前一章中多次讨论过的那样。在这个例子中，$M=N=480$。我们知道，高斯形函数没有零值，因此这里不用关心它。然而，尽管如此，退化值变得非常小，以至于全逆滤波的结果［见图 5.27(a)］毫无用处。产生这个失败结果的原因如式(5.7-2)所讨论的那样。



a b c d

图 5.27　用式(5.7-1)复原图 5.25(b)：(a)用全滤波器的结果；(b)截止半径为 40 时用 H 的结果；(c)半径为 70 时的结果；(d)半径为 85 时的结果

图5.27(b)到图5.27(d)显示了 $G(u, v)/H(u, v)$的截止值的结果，半径分别为 40, 70 和 85。截止是通过对这个比值应用一个阶数为 10 的布特沃斯低通函数来实现的。这在期望的半径处提供了较尖锐(但平滑)的过渡。半径在 70 左右时视觉效果最好［见图5.27(c)］，半径在 70 以下时图像开始变得模糊，如图5.27(b)所示，它是用半径 40 得到的。值大于 70 时，图像开始退化，如图5.27(d)所示，它是用半径 85 得到的。在噪声“窗帘”后面的图像中，图像的内容几乎是可见的，但噪声明显地支配着该结果。进一步增大半径值，会使得图像越来越像图5.27(a)。

373 ~ 374

这个例子的结果表明，一般直接逆滤波的性能是较差的。接下来的三节将探讨怎样改进直接逆滤波的性能。

5.8 最小均方误差(维纳)滤波

前一节讨论的逆滤波方法并未清楚地说明怎样处理噪声。本节讨论一种综合退化函数和噪声统计特征进行复原处理的方法。该方法建立在图像和噪声都是随机变量的基础上，目标是找到未污染图像 f 的一个估计 $\hat{f}$，使它们之间的均方误差最小。这种误差度量由下式给出：

> 注意整个图像都被考虑为随机变量，详见2.6.8节末尾的讨论。

$$e^2 = E\left\{(f-\hat{f})^2\right\} \tag{5.8-1}$$

式中，$E\{\cdot\}$是参数的期望值。这里假设噪声和图像不相关，其中一个或另一个有零均值，且估计中的灰度级是退化图像中灰度级的线性函数。基于这些条件，式(5.8-1)中误差函数的最小值在频率域中由如下表达式给出：

$$\begin{aligned}\hat{F}(u,v) &= \left[\frac{H^*(u,v)S_f(u,v)}{S_f(u,v)|H(u,v)|^2+S_\eta(u,v)}\right]G(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)|^2+S_\eta(u,v)/S_f(u,v)}\right]G(u,v) \\ &= \left[\frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2+S_\eta(u,v)/S_f(u,v)}\right]G(u,v)\end{aligned} \tag{5.8-2}$$

这里，我们用到了这样一个事实：一个复数量与其共轭的乘积等于该复数量幅度的平方。这个结果就是众所周知的维纳滤波，该概念由N. Wiener[1942]首次提出。由方括号中的项组成的滤波器通常也称为*最小均方误差滤波器*或*最小二乘误差滤波器*。本章最后列出了一些参考资料，包括了维纳滤波器的详细推导过程。注意式(5.8-2)的第一行，维纳滤波器没有逆滤波中退化函数为零的问题，除非对于相同的 u 值和 v 值，整个分母都是零。

式(5.8-2)中各项的意义如下所示：

$H(u, v)$ = 退化函数

$H^*(u, v) = H(u, v)$的复共轭

$|H(u, v)|^2 = H^*(u, v)H(u, v)$

$S_\eta(u, v) = |N(u, v)|^2$ = 噪声的功率谱［见式(4.6-18)］[①]

$S_f(u, v) = |F(u, v)|^2$ = 未退化图像的功率谱

与前面一样，$H(u, v)$是退化函数的变换，而$G(u, v)$是退化图像的变换。空间域中复原的图像由频率域估计 $\hat{F}(u,v)$ 的傅里叶反变换给出。注意，如果噪声为零，则噪声功率谱消失，并且维纳滤波简化为逆滤波。

许多有用的度量是以噪声和未退化图像的功率谱为基础的，其中最重要的一种是信噪比，它在频率域中用下式来近似：

$$\mathrm{SNR} = \frac{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}|F(u,v)|^2}{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}|N(u,v)|^2} \tag{5.8-3}$$

374 ~ 376

① 项$|N(u, v)|^2$也称为噪声的自相关。该术语源自相关定理(见表4.3中的性质7)。当两个函数相同时，两个函数的相关就变成了自相关，该性质的右侧变成 $N^*(u, v)N(u, v)$，它等于$|F(u, v)|^2$。类似的说明同样适用于$|F(u, v)|^2$，即它是图像的自相关。第12章中将详细探讨相关。

该比值给出了携带信息的信号功率（即原始的或退化的原图像）水平与噪声功率水平的度量。携带低噪声的图像有较高的 SNR，而携带较高噪声的同一幅图像有较低的 SNR。这一比值是一个有限值，但在用于表征复原算法的性能时，它是一个重要的度量。

式(5.8-1)中以统计形式给出的均方误差，也可描述为涉及原图像和复原图像的和的形式：

$$\mathrm{MSE} = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[f(x,y)-\hat{f}(x,y)\right]^2 \tag{5.8-4}$$

事实上，如果把复原图像考虑为“信号”，而把复原图像和原图像的差考虑为噪声，那么我们可以将空间域中的信噪比定义为

$$\mathrm{SNR} = \frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\hat{f}(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[f(x,y)-\hat{f}(x,y)\right]^2} \tag{5.8-5}$$

f 和 $\hat{f}$ 越接近，该比值就较大。有时，用这一度量的均方根来代替，这种情况下，该比值称为均方根信噪比或均方根误差。正如以前我们多次提到的那样，必须记住该定量度量与感觉的图像质量并没有很好的必然关系。

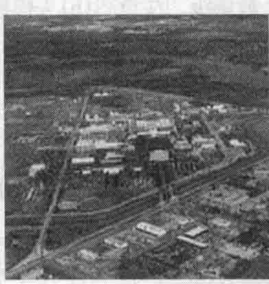
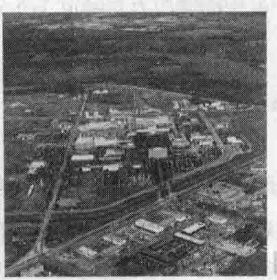
当我们处理白噪声时，谱$|N(u,v)|^2$是一个常数，这大大简化了处理。然而，未退化图像的功率谱很少是已知的。当这些量未知或不能估计时，经常使用的一种方法是由下面的表达式来近似式(5.8-2)：

$$\hat{F}(u,v) = \left[\frac{1}{H(u,v)}\frac{|H(u,v)|^2}{|H(u,v)|^2+K}\right]G(u,v) \tag{5.8-6}$$

式中，K 是一个加到$|H(u,v)|^2$的所有项上的特定常数。下面的几个例子说明了这一表达式的应用。

例 5.12　逆滤波和维纳滤波的比较。

图5.28说明了维纳滤波与直接逆滤波相比的优点。图5.28(a)是图5.27(a)的全逆滤波的结果。类似地，图5.28(b)是图5.27(c)的半径受限的逆滤波结果。为便于比较，这里重新列出了这些图像。图5.28(c)显示了使用式(5.8-6)通过例5.11中使用的退化函数得到的结果。K 值的交互式选择是为了找到最好的视觉效果。在这个例子中，与直接逆滤波方法相比，维纳滤波的效果显然更好。比较图 5.25(a)和图 5.28(c)，我们会发现维纳滤波的结果非常接近原图像的外观。

a b c

图 5.28　逆滤波和维纳滤波的比较：(a)图 5.25(b)的全逆滤波结果；(b)半径受限的逆滤波结果；(c)维纳滤波的结果

例 5.13　维纳滤波的进一步比较。

图5.29 的第一行从左到右分别显示了被均值为零、方差为 650 的加性高斯噪声严重污染的图5.26(b)的模糊图像，直接逆滤波的结果，以及维纳滤波的结果。用式(5.8-6)的维纳滤波，通过例 5.10 的 $H(u, v)$

376～377

和交互地选择 K，可以得到最好的视觉效果。如所期望的那样，逆滤波产生了一幅不太适用的图像。注意，逆滤波后的图像中的噪声非常强，以至于它的结构就在去模糊滤波器的方向上。维纳滤波的结果并不完美，但它的确给了我们一些关于图像内容的线索。要辨认图中的文字，还是有点难度。

图5.29的第二行显示了同样的序列，但噪声方差的水平降低了一个数量级。这种降低对逆滤波没有多少影响，但维纳滤波的效果明显改进了，图中的文字容易辨认多了。图5.29的第三行相对于第一行噪声方差降低了 5 个数量级。事实上，图5.29(g)中的图像已没有可见的噪声。在这种情况下，逆滤波的结果相当有趣。噪声仍很明显，但是文字能通过噪声“窗帘”看见。这是关于式(5.7-2)的一个很好的例子。换句话说，正如从图5.29(h)中明显看到的那样，逆滤波消除图像模糊的能力是很强的。然而，噪声还是支配着结果。即使我们能看到图5.29(b)和图5.29(e)中噪声背后的内容，但字符还是显得有点模糊。图5.29(i)中的维纳滤波结果是最好的，视觉上非常接近于图5.29(a)中的原图像。只要退化函数的合理估计是可用的，这些类型的结果就是使用维纳滤波可能得到的结果。

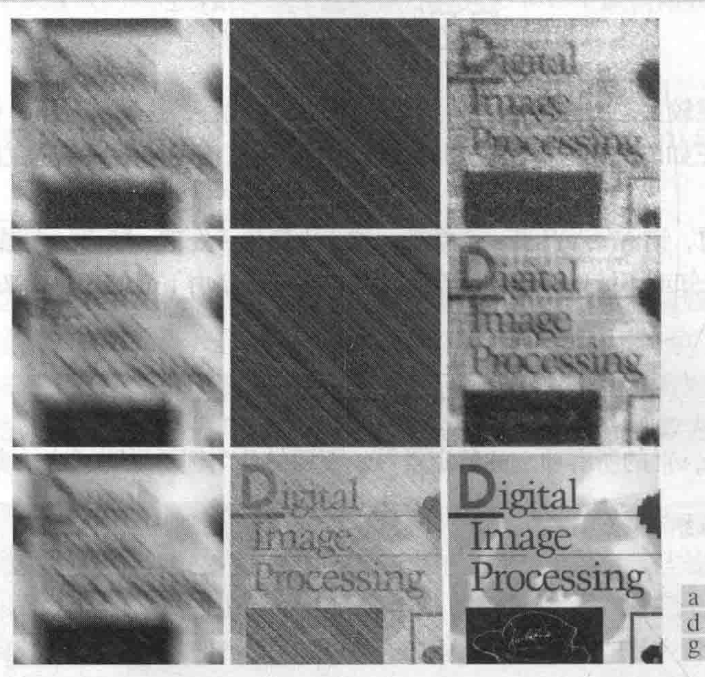


图 5.29 (a)由运动模糊和加性噪声污染的 8 比特图像；(b)逆滤波的结果；(c)维纳滤波的结果；(d)~(f)同样的图像序列，但噪声方差降低了一个数量级；(g)~(i)同样的图像序列，但噪声方差比图(a)中的噪声方差降低了 5 个数量级。注意图(h)去模糊图像透过噪声“窗帘”清晰可见

5.9 约束最小二乘方滤波

关于退化函数 H 的某些知识必须已知的问题，对于本章讨论的所有方法很常见。然而，维纳滤波却存在一些其他的困难：未退化图像和噪声的功率谱必须是已知的。前面的章节表明，使用式(5.8-6)给出的近似方法能得到很好的结果。然而，功率谱比的常数估计并不总是一个合适的解。

本节讨论的方法仅要求噪声方差和均值的知识。正如5.2.4节讨论的那样，这些参数通常可从一幅给定的退化图像计算出来，因此这是一个很重要的优点。另一个不同是，维纳滤波建立在最小化统计准则的基础之上，因此在平均意义上它是最优的。本节给出的算法具有显著的特点，即对于其应用到的每幅图像都能产生最优结果，当然，这些理论上满足的最优准则与动态的视觉感知并没有关系。因此，算法的选择往往由结果图像的视觉感知质量决定(至少部分是这样的)。

377 ~ 379

通过使用式(4.6-23)给出的卷积定义，并像 2.6.6 节解释的那样，我们可将式(5.5-16)表达为如下向量-矩阵形式：

关于向量和矩阵的简要回顾，请查阅本书的网站。

$$\boldsymbol{g} = \boldsymbol{Hf} + \boldsymbol{\eta} \tag{5.9-1}$$

例如，假设 $g(x, y)$ 的大小为 $M \times N$。然后，我们可以使用 $g(x, y)$ 的第一行中的图像元素构成向量 $\boldsymbol{g}$ 的第一组 N 个元素，用第二行构成下一组 N 个元素，以此类推。结果向量将有 $MN \times 1$ 维。这也同样是 $\boldsymbol{f}$ 和 $\boldsymbol{\eta}$ 的维数，因为这些向量是以同样的方式构成的。矩阵 $\boldsymbol{H}$ 有 $MN \times MN$ 维，其元素由式(4.6-23)中的卷积元给出。

把图像复原问题简化为简单的矩阵操作的结论看起来很合理。遗憾的是，事实上并非如此。例如，假设我们正在处理一幅中等大小的图像，譬如大小为 $M = N = 512$，那么式(5.9-1)中的向量为 262 144×1 维，矩阵 $\boldsymbol{H}$ 为 262 144×262 144 维。计算这种大小的向量和矩阵并不是一件简单的任务。由于 $\boldsymbol{H}$ 对噪声高度敏感，因此问题被进一步复杂化(在前面两节中我们已了解到噪声的影响，因此对此不应感到奇怪)。然而，明确地以矩阵形式来表达复原问题确实可以简化复原技术的推导。

尽管我们不能完全推导已存在的约束最小二乘法，但这种方法在矩阵表示法中的确有其根源。在本章的最后会给出参考资料，涵盖详细推导的来源。该方法的核心是 $\boldsymbol{H}$ 对噪声的敏感性问题。减少噪声敏感性问题的一种方法是以平滑度量的最佳复原为基础的，如一幅图像的二阶导数（“拉普拉斯变换”这位“老朋友”）。意味深长的是，必须用手边问题的参数来约束复原。因此，我们的期望是找一个最小准则函数 C，定义如下：

使用代数技术来进行图像复原的整章内容，请查阅本书网站上的 Tutorials 部分。

$$C = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left[\nabla^2 f(x, y) \right]^2 \tag{5.9-2}$$

其约束为

$$\left\| \boldsymbol{g} - \boldsymbol{H}\hat{\boldsymbol{f}} \right\|^2 = \left\| \boldsymbol{\eta} \right\|^2 \tag{5.9-3}$$

式中，$\|\boldsymbol{w}\|^2 \triangleq \boldsymbol{w}^{\mathrm{T}} \boldsymbol{w}$ 是欧几里得向量范数[①]，$\hat{\boldsymbol{f}}$ 是未退化图像的估计。拉普拉斯算子 ∇^2 在式(3.6-3)中定义过。

这个最佳化问题在频率域中的解决由下面的表达式给出：

$$\hat{F}(u, v) = \left[\frac{H^*(u, v)}{\left|H(u, v)\right|^2 + \gamma \left|P(u, v)\right|^2} \right] G(u, v) \tag{5.9-4}$$

式中，γ 是一个参数，必须对它进行调整以满足式(5.9-3)的条件，$P(u, v)$ 是函数

$$p(x, y) = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{5.9-5}$$

的傅里叶变换。在3.6.2节中介绍拉普拉斯算子时我们曾见过这个函数。如早些时候注意到的那样，记住，$p(x, y)$ 和所有其他相关的空间域函数很重要，在用式(5.9-4)计算其傅里叶变换之前要用零进行适当的填充，正如4.6.6节所讨论的那样。注意，当 γ 为 0 时，式(5.9-4)简化为逆滤波。

379 ~ 380

① 回忆可知，对于有着 n 个元素的向量 $\boldsymbol{w}$，有 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} = \sum_{k=1}^{n} w_k^2$，其中 w_k 是 $\boldsymbol{w}$ 的第 k 个元素。

例 5.14 维纳滤波和最小二乘方滤波的比较。

图5.30显示了使用约束最小二乘方滤波处理图5.29(a)、图5.29(d)和图5.29(g)的结果，为产生最好的视觉效果，γ 的值是手工选择的。使用同样的过程，产生的维纳滤波结果如图5.29(c)、图5.29(f)和图5.29(i)所示。比较约束最小二乘方滤波和维纳滤波的结果，发现前者对于高噪声和中等噪声的情况，产生的结果要稍好一些，而对于低噪声情况，两种滤波产生基本相等的结果。当手工选择参数以取得更好的视觉效果时，约束最小二乘方滤波的效果有可能比维纳滤波的效果更好。式(5.9-4)中的参数 γ 是一个标量，而式(5.8-6)中的 K 值是两个未知频率域函数的近似比，该比值很少是常数。因此，手工选择 γ 将得出未退化图像的更为准确的估计。



图 5.30 约束最小二乘方滤波的结果。请将图 5.29(c)、图 5.29(f)和图 5.29(i)所示的维纳滤波结果分别与图(a)、图(b)和图(c)比较

如前面的例子所示，交互地调整参数 γ 达到可接受的结果是可能的。然而，如果我们对最优化感兴趣，就必须调整参数 γ 以满足式(5.9-3)中的约束条件。迭代计算 γ 的过程如下。

定义一个“残差”向量 $\boldsymbol{r}$ 为

$$\boldsymbol{r} = \boldsymbol{g} - \boldsymbol{H}\hat{\boldsymbol{f}} \tag{5.9-6}$$

因为，由式(5.9-4)的解可知，$\hat{F}(u,v)$（即隐含的 $\hat{f}$）是 γ 的函数，所以 $\boldsymbol{r}$ 也是该参数的函数。可以证明(Hunt[1973])

$$\phi(\gamma) = \boldsymbol{r}^{\mathrm{T}}\boldsymbol{r} = \|\boldsymbol{r}\|^2 \tag{5.9-7}$$

是 γ 的单调递增函数。我们要做的是调整 γ 使得

$$\|\boldsymbol{r}\|^2 = \|\eta\|^2 \pm a \tag{5.9-8}$$

其中，a 是一个精确度因子。在式(5.9-6)看来，如果 $\|\boldsymbol{r}\|^2 = \|\boldsymbol{\eta}\|^2$，那么严格满足式(5.9–3)的约束条件。因为 $\phi(\gamma)$ 是单调的，所以寻找满足要求的 γ 值并不困难。一种方法是

1. 指定 γ 的一个初始值。
2. 计算 $\|\boldsymbol{r}\|^2$。
3. 若满足式(5.9-8)，则停止；否则，若 $\|\boldsymbol{r}\|^2 < \|\boldsymbol{\eta}\|^2 - a$，则增大 γ，若 $\|\boldsymbol{r}\|^2 > \|\boldsymbol{\eta}\|^2 + a$，则减小 γ，然后返回步骤 2。使用 γ 的新值，通过式(5.9-4)重新计算最佳估计 $\hat{F}(u,v)$。

其他的方法，譬如牛顿-拉弗森算法，也可用于改进收敛的速度。

为了使用这一算法，我们需要 $\|\boldsymbol{r}\|^2$ 和 $\|\boldsymbol{\eta}\|^2$ 的值。为了计算 $\|\boldsymbol{r}\|^2$，由式(5.9-6)得

$$R(u,v) = G(u,v) - H(u,v)\hat{F}(u,v) \tag{5.9-9}$$

由此，我们可以通过计算 $R(u, v)$的傅里叶反变换得到 $r(x, y)$。然后有

381 ~ 382

$$\|\boldsymbol{r}\|^2 = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} r^2(x,y) \tag{5.9-10}$$

计算$\|\boldsymbol{\eta}\|^2$可以得出一个重要的结果。首先，考虑整幅图像上的噪声方差，该方差可以使用取样平均的方法来估计，如 3.3.4 节讨论的那样：

$$\sigma_\eta^2 = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\eta(x,y) - m_\eta\right]^2 \tag{5.9-11}$$

式中，

$$m_\eta = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\eta(x,y) \tag{5.9-12}$$

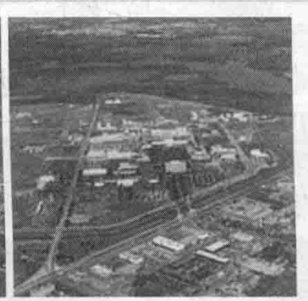
是样本均值。参考式(5.9-10)的形式，我们注意到式(5.9-11)中的双重求和等于$\|\boldsymbol{\eta}\|^2$。这就给出了如下表达式：

$$\|\boldsymbol{\eta}\|^2 = MN\left[\sigma_\eta^2 + m_\eta^2\right] \tag{5.9-13}$$

这是非常有用的结果。它告诉我们，仅用噪声均值和方差的知识，我们就可以实现一个最佳复原算法。这些值并不难估计(见5.2.4节)，但要假设噪声和图像灰度值不相关。这是本章讨论的所有方法的一个基本假设。

例 5.15　最佳约束最小二乘方滤波的迭代估计。

图 5.31(a)显示了使用刚才描述的算法来估计用于复原图 5.25(b)的最佳滤波得到的结果。γ 的初始值是 10^{-5}，调整 γ 的校正因子是 10^{-6}，a 的值为 0.25。用于生成图5.25(a)的指定噪声参数相同：噪声方差为 10^{-5}，均值为 0。复原结果几乎与图5.28(c)一样好，图5.28(c)是由维纳滤波得到的结果，为得到最好的视觉效果，手工指定了 K 值。图5.31(b)显示了使用噪声参数的错误估计值时发生的情况。在这种情况下，指定的噪声方差为 10^{-2}，并且均值保留为 0。在这种情况下，结果有相当大的模糊。

a b

图 5.31　(a)对图 5.16(b)使用正确的噪声参数迭代确定的约束最小二乘方复原；(b)使用错误的噪声参数得到的结果

正如本节开始指出的那样，约束最小二乘方意义下的最佳复原在视觉效果上并不意味着最好。根据退化和噪声的性质及大小，算法中交互确定最佳估计的其他参数，在最终结果中也起着很重要的作用。通常，自动确定的复原滤波器比人为调整滤波器参数的复原结果要差。特别是约束最小二乘方滤波器完全由单一的标量参数来决定时更是如此。

5.10　几何均值滤波

我们可对 5.8 节中讨论的维纳滤波器稍加推广。这种推广就是所谓的几何均值滤波器的形式：

$$\hat{F}(u,v) = \left[\frac{H^*(u,v)}{|H(u,v)|^2}\right]^\alpha \left[\frac{H^*(u,v)}{|H(u,v)|^2 + \beta\left[\dfrac{S_\eta(u,v)}{S_f(u,v)}\right]}\right]^{1-\alpha} G(u,v) \tag{5.10-1}$$

式中，α 和 β 是正的实常数。几何均值滤波器由两个括号内的幂次分别为 α 和 $1-\alpha$ 的表达式组成。

382
~
383

当$\alpha=1$时，该滤波器退化为逆滤波器；当$\alpha=0$时，该滤波器变为所谓的参数维纳滤波器，参数维纳滤波器在$\beta=1$时还原为标准的维纳滤波器。如果$\alpha=1/2$，则滤波器变成相同幂次的两个量的积，这是几何均值的定义，这样就给出了这种滤波器的命名。当$\beta=1$时，随着α减小到1/2以下，滤波器的性能越来越接近逆滤波器。类似地，当α增大到1/2以上时，滤波器更接近维纳滤波器。当$\alpha=1/2$且$\beta=1$时，该滤波器通常也称为谱均衡滤波器。式(5.10-1)在实现复原滤波时非常有用，因为它表示了合并为单个表达式的滤波器族。

5.11 由投影重建图像

在本章的前几节中，我们讨论了各种复原退化图像的技术。本节研究由一系列投影重建一幅图像的问题，集中讨论X射线计算机断层(CT)。这是最早且应用最广泛的CT类型，也是当前数字图像处理在医学中的主要应用之一。

如第1章注释的那样，通常也用术语计算机轴向断层术(CAT)来表示CT。

5.11.1 引言

重建问题在原理上很简单，并且可以直接用直观的定性方法加以描述。首先，考虑图5.32(a)，它由均匀背景上的单一物体组成。它的物理意义可解释如下：假设该图像是人体三维区域的一个横断面。还假设图像中的背景表示均匀的软组织，它所环绕的物体是一个肿瘤，该肿瘤也是均匀的，但有较高的吸收特性。

接着，如图5.32(a)所示，假设我们用一束细的、平行的X射线从左到右扫描(通过图像平面)，并且假设物体吸收的射线束能量比背景吸收的射线束能量多，这是典型的情况。利用放在该区域另一端的X射线吸收检测器带将产生所示的信号(吸收断面图)。该信号的幅度(亮度)与吸收成正比①。我们能够观察到信号中的任意一点都是穿过相应空间点的该射线束中单一射线吸收值的和(这样的和通常称为射线和)。这时，有关物体的所有信息就是这个一维的吸收信号。

由单个投影，我们无法确定沿射线路径处理的是单个物体还是多个物体，但我们可以仅基于这一信息来创造一幅图像的方法开始重建。如图5.32(b)所示，该方法是沿着射线来的方向把一维信号反投影回去。穿过二维区域的一维信号的反投影过程，有时可想象为把投影穿过该区域反“涂抹”回去。在数字图像中，这意味着沿垂直于射线方向复制横穿图像的相同一维信号。例如，图5.32(b)就是由复制矩形图像的所有列中的一维信号创建的。显然，刚刚描述的方法就称为反投影法。

接着，假设我们把源-检测器对的位置旋转90°，如图5.32(c)所示。重复前段解释的步骤，在垂直方向生成一幅反投影图像，如图5.32(d)所示。我们把这一结果加到前面的反投影图像上继续重建工作，结果如图5.32(e)所示。现在，我们可以说我们感兴趣的物体已包含在所示的方形中，其幅度是单个投影幅度的两倍。稍加思考后，就会了解使用刚刚讨论的方法得到更多的视图，我们应能得到关于物体形状的更多信息。事实上，如图5.33所示，所发生的情况说明这是正确的。随着投影数量的增加，不相交反投影的强度相对于多个反投影相交区域的强度会降低。最终结果是，较亮区域将支配结果，当为显示而调节亮度时，很少相交或不相交的反投影将减弱为背景。

由32个投影形成的图5.33(f)说明了这一概念。但要注意，尽管这幅重建的图像是对原始物体形状的较好近似，但它却被“晕环”效应所模糊，关于该信息可在图5.33中的渐进步骤中看到。例

384
~
385

① 关于X射线源和检测器的物理知识超出了我们探讨的范围，我们的讨论主要是CT的图像处理。关于X射线图像形成的物理原理，请参阅 Prince and Links[2006]。

如，图5.33(e)中的"晕环"以星状形式出现，其亮度比物体低，但又比背景高。当视图数量增加时，"晕环"的形状变成一个圆，如图5.33(f)所示。在 CT 重建中，模糊是一个重要问题，其解决办法将在 5.11.5 节讨论。最后，从图5.32 和图 5.33 的讨论，我们得出结论：相隔 180°的投影互为镜像图像，因此，为了产生重建所要求的所有投影，我们只要按圆周的一半角度增量来考虑即可。

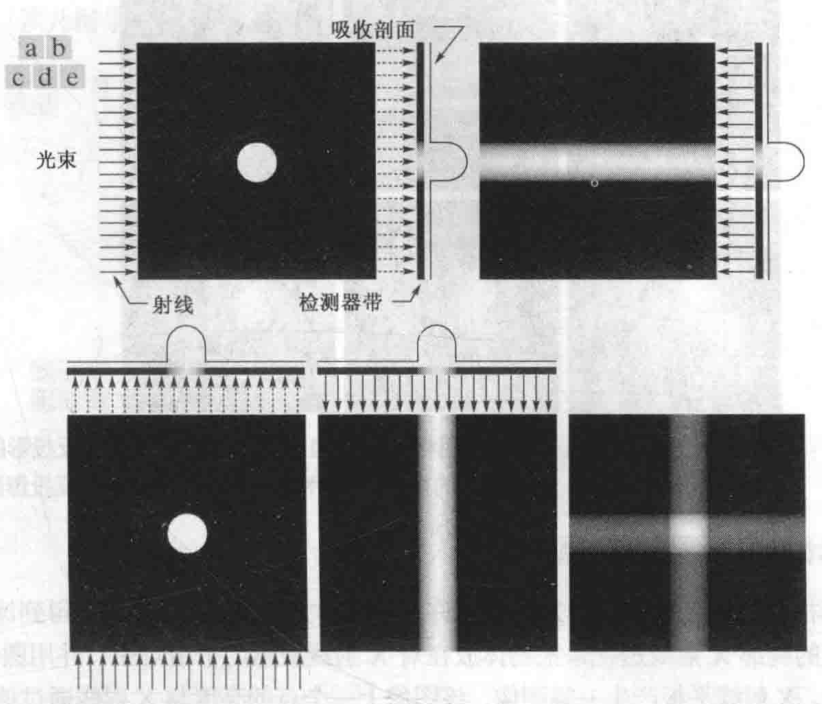


图 5.32　(a)显示一个简单物体的平坦区域，一个输入的平行射线束，一个检测器带；(b)感知条带数据的反投影的结果(即一维吸收剖面)；(c)旋转 90°后的射线束和检测器；(d)反投影；(e)图(b)和图(d)的和。反投影相交处的亮度是各个反投影的亮度的两倍

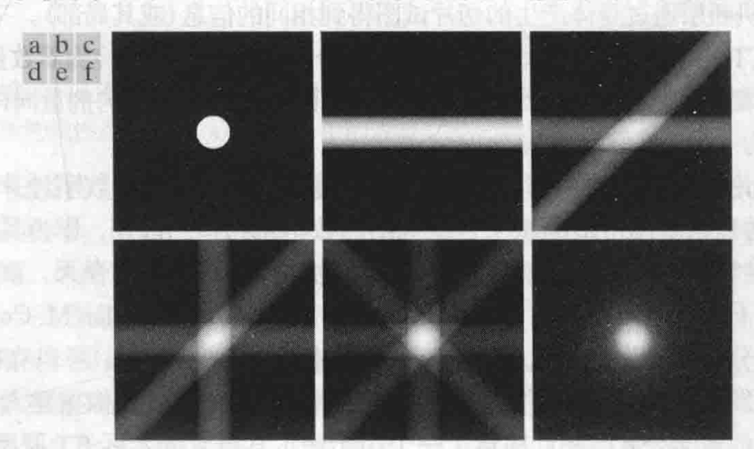


图 5.33　(a)与图 5.32(a)相同的图像；(b)~(e)使用间隔为 45°的 1、2、3、4 个反投影的重建；(f)间隔 5.625°的 32 个反投影的重建(注意模糊效果)

例 5.16　包含两个物体的简单平面的反投影。

图 5.34 说明了稍微复杂一些的区域的反投影法重建，该区域包含两个具有不同吸收特性的物体。图 5.34(b)显示了用一个反投影的结果。我们注意到该图的三个主要特性，从下到上：一个细的水平灰色带与小物体还不能最后断定的部分相对应；在它上边较亮的条带(吸收更多)对应于由两个物体共享的区域；以及对应于椭圆物体剩余部分的上边的条带。图 5.34(c)和图 5.34(d)分别显示了使用相隔 90°的两个投

385~386

影和相隔 45°的 4 个投影的重建结果。这些图形的解释类似于对图5.33(c)到图5.33(e)的讨论。图5.34(e)和(f)显示了分别使用 32 个和 64 个反投影的更精确重建结果。这两个结果在视觉上非常接近，而且两者都显示了之前提到的模糊问题，模糊问题的解决办法将在5.11.5节中讨论。

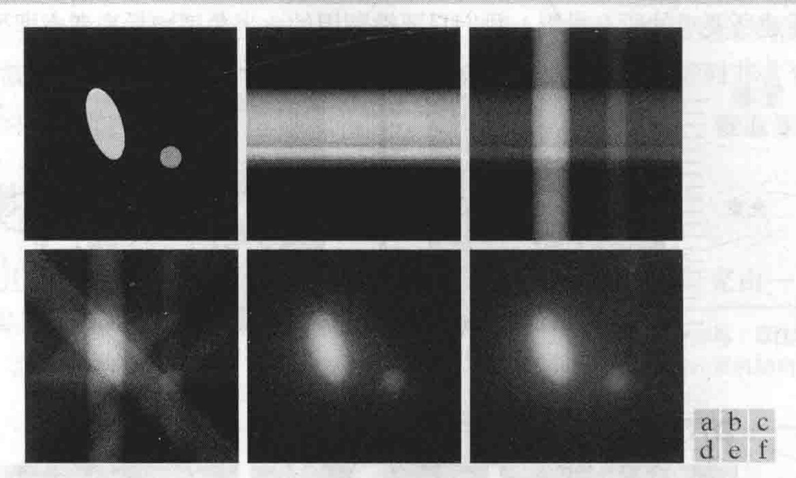


图 5.34 (a)有两个物体的区域；(b)~(d)用相隔 90°的 1、2 和相隔 45°的 4 个反投影的重建；(e)用相隔 5.625°的 32 个反投影的重建；(f)用相隔 2.8125°的 64 个反投影的重建

5.11.2 计算机断层(CT)原理

X 射线计算机断层的目的是使用 X 射线从许多不同的方向穿过物体，进而得到该物体内部结构的三维描述。传统的胸部 X 射线透视是把物体放在对 X 射线敏感的平板对面，并用圆锥形 X 射线束照射这个体得到的。X 射线平板产生一幅图像，该图像上一个点的亮度与 X 射线通过该物体后照射到该点上的 X 射线能量成正比。这幅图像就是在前节中讨论过的投影的二维图像。我们可以对整幅图像进行反投影，并创建一个三维物体。通过在多个角度重复该过程，并把反投影相加，就可产生胸腔结构的三维再现。计算机断层通过身体产生的切片试图得到相同的信息(或其局部)。三维描述可以通过堆积这些切片得到。CT 的实现要经济得多，因为得到高分辨率切片所需的检测器数量，要比产生相同分辨率的一个完整二维投影所需的检测器数量少得多。计算负担和 X 射线的剂量同样要降低，这样才能使一维投影 CT 成为一种更实际的方法。

如上一章中讨论傅里叶变换时那样，CT 所要求的基本数学概念在数字计算机能够实用之前已存在很多年。CT 的理论基础可追溯至来自维也纳的数学家约翰·雷登，作为其线积分工作的一部分，他于 1917 年推导了一种沿平行射线对一个二维物体投影的方法。今天，该方法就是通常所指的雷登变换，这是下一节讨论的主题。45 年后，埃兰·M·考玛克(Allan M. Cormack)——塔夫茨大学的物理学者部分地重新发现了这些概念，并把它们应用到 CT 上。考玛克在 1963 年和 1964 年发表了他的最初发现，并且说明了如何从不同角度得到的 X 射线图像重建人体横截面图像。他给出了重建所需要的数学公式，并且构建了一个用于展示其概念的实际 CT 原型。同时，伦敦 EMI 公司的电子工程师高德弗里·N·豪斯菲尔德(Godfrey N. Hounsfield)及其同事也独立给出了类似的解决方法，并建立了第一台医学 CT 机。由于他们对医学断层的贡献，考玛克和豪斯菲尔德共同获得了 1979 年的诺贝尔医学奖。

第一代(G1)CT 扫描仪采用“铅笔”型 X 射线束和单个检测器，如图 5.35(a)所示。对于一个给定的旋转角度，源-检测器对沿着所示的线性方向增量式地平移。投影(类似于图 5.32)由测量每个增量平移处检测器的输出产生。完成线性平移之后，旋转源-检测器组合，并重复该过程得到不同角度上的另一个投影。该过程在 0°到 180°内对所有期望的角度重复，生成一组完整的投影，就

386~388

像前一节解释的那样，用反投影产生一幅图像。物体顶部的十字标记表明与源-检测器对构成的平面相垂直的方向的运动。增量式地将物体移过源-检测器平面(每完成一次扫描后)，产生一组横截面图像(切片)。堆积这些图像就可生成人体截面的三维体。医学成像的第一代扫描仪不久就不再生产，但由于它们可产生平行的射线束(如图5.32 所示)，它们的几何学原理作为介绍 CT 成像的基础是有优势的。正如下一节要讨论的那样，该几何学原理是推导由投影实现图像重建所需公式的起点。

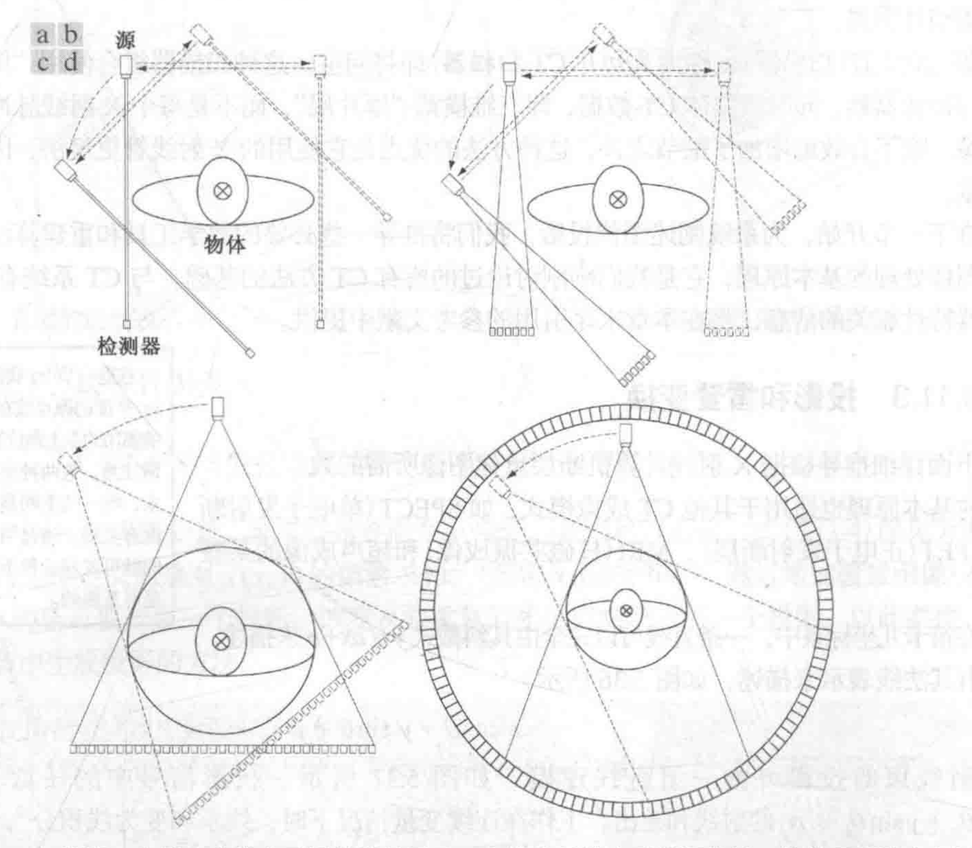


图 5.35　四代 CT 扫描器。点箭头直线表示增量线性运动。点箭头弧线表示增量旋转。物体顶部的十字标记表示垂直于纸面的线性运动。图(a)和图(b)中的双箭头表示源/检测器单元被平移后又回到了其原始位置

第二代(G2)CT 扫描器［见图5.35(b)］与第一代扫描器的工作原理相同，但所用射线束是扇形的。这就允许使用多个检测器，因而源-检测器对的平移较少。第三代(G3)扫描器较前两代 CT 在几何原理上有较大的改进。如图 5.35(c)所示，G3 扫描器使用足够长的一族检测器(约有 1000 个独立的检测器)来覆盖一个更宽射线束的整个视野。因此，每个角度的增量都会产生一个完整的投影，从而消除了如 G1 和 G2 扫描器那样对源-检测器对的平移的需求。第四代(G4)扫描器更进一步。它使用一个圆环检测器(约有 5000 个独立的检测器)，仅射线源旋转。G3 和 G4 扫描器的重要优点是速度快，主要缺点是造价高和较大的 X 射线散射，它需要比 G1 和 G2 扫描器更高的剂量才能达到可比拟的信噪比特性。

新的扫描模式正在开始使用。例如，第五代(G5)CT 扫描器，即所知道的电子束计算机断层(EBCT)扫描器，取消了所有的机械运动，而改为以电磁方式来控制电子束。通过触发环绕病人的钨极板，这些电子束产生 X 射线，然后 X 射线被整形为通过病人的扇形射线束，并激发如 G4 扫描器那样的检测器环。

传统方法得到的 CT 图像在扫描期间要求病人保持不动才能生成图像。然后，病人在垂直于成像平面上用机动工作台递增移动位置时，停止扫描。再后，得到下一幅图像；重复该过程多次，直

388
~
389

到覆盖人体某个特定的截面所要求的数量为止。虽然不到一秒就可得到一幅图像，但在图像获取期间要求病人摒住呼吸(如腹部和胸部扫描)。完成 30 幅图像的采集过程可能需要几分钟时间。针对该问题所用的一种方法是螺旋 CT，有时也称为第六代(G6) CT。在该方法中，G3 和 G4 扫描器使用一种所谓的滑动环来配置，它不需要源-检测器和处理单元之间的电气和信号连接。然后，源-检测器对连续旋转 360°，同时病人在垂直于扫描的方向恒速移动。结果是连续的螺旋数值，这些数据经处理后就可得到各幅切片图像。

第七代(G7)扫描器(也称为*多切片 CT 扫描器*)即将问世，这种扫描器组合使用“厚”扇形射线束和平行检测器族，同时收集体 CT 数据，即三维横截“厚片层”，而不是每个 X 射线脉冲产生单个横截面图像。除了有效地增加了细节之外，这种方法的优点是它使用的 X 射线管更经济，因此降低了成本和剂量。

在下一节开始，为系统阐述图像投影，我们将推导一些必要的数学工具和重建算法。我们的关注点是图像处理的基本原理，它是我们刚刚讨论过的所有 CT 方法的基础。与 CT 系统有关的机械和源-检测器特性相关的信息，将在本章末尾引用的参考文献中提供。

5.11.3 投影和雷登变换

在这一节中，我们遵从 CT 约定并将 xy 平面的原点放在中心位置而不是之前惯用的左上角位置(见 2.4.2 节)。但请注意，这两种坐标系均是右手坐标系，唯一的不同是我们的图像坐标系没有负轴。通过简单地移动原点，我们就可发现这种不同，因此两种表示是可互换的。

下面详细推导根据 X 射线计算机断层重建图像所需的数学公式，相同的基本原理也适用于其他 CT 成像模式，如 SPECT(单电子发射断层)、PET(正电子发射断层)、MRI(核磁共振成像)和超声成像的某些模式。

在笛卡儿坐标系中，一条直线可以完全由其斜截式 $y = ax+b$ 来描述，或可由其法线表示来描述，如图 5.36 所示：

$$x\cos\theta + y\sin\theta = \rho \tag{5.11-1}$$

平行射线束的投影可由一组直线建模，如图 5.37 所示。投影信号中的任意一点由沿直线 $x\cos\theta_k + y\sin\theta_k = \rho_j$ 的射线和给出。工作在连续变量情况下时，线求和变为线积分[①]，由下式给出：

$$g(\rho_j, \theta_k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta_k + y\sin\theta_k - \rho_j)\,\mathrm{d}x\,\mathrm{d}y \tag{5.11-2}$$

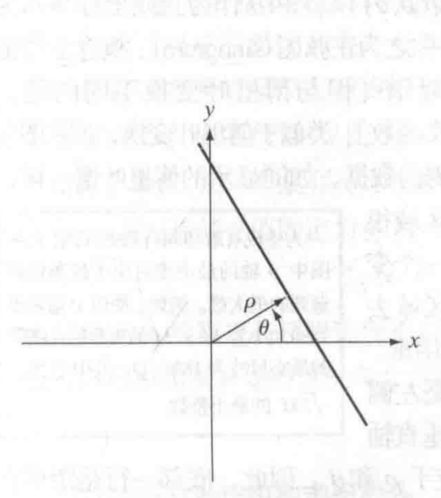
式中，我们使用了 4.5.1 节中讨论的冲激函数 δ 的性质。换句话说，除非 δ 的参量是零，否则式(5.11-2)的右边为零，它指出积分只沿线 $x\cos\theta_k + y\sin\theta_k = \rho_j$ 计算。如果考虑 ρ 和 θ 的所有值，那么前一公式推广为

$$g(\rho,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)\,\mathrm{d}x\,\mathrm{d}y \tag{5.11-3}$$

给出沿 xy 平面中任意一条线的 $f(x,y)$ 的投影(线积分)的公式，就是前一节提到的雷登变换。符号 $\mathcal{R}\{f(x,y)\}$ 或 $\mathcal{R}\{f\}$ 有时用于代替式(5.11-3)中的 $g(\rho,\theta)$ 来表示 f 的雷登变换，但式(5.11-3)中所用的符号类型更为常见。随后的讨论更明显，雷登变换是投影重建的基石，计算机断层是其在图像处理领域的主要应用。

① 在第 4 章中，我们使用 (t,z) 来表示连续图像坐标，而用 (x,y) 来表示离散坐标。当时，这种区别很重要，因为我们在推导从连续到取样后的量的基本概念。在当前的讨论中，由于我们会在连续与离散坐标之间多次往返，坚持这种约定很可能会导致不必要的混淆。因此，为遵从已出版文献中的约定(见 Prince and Links[2006])，我们会根据上下文来确定坐标 (x,y) 是连续的还是离散的。当坐标是连续坐标时，我们会看到积分，而坐标为离散坐标时，会看到求和。

390 ~ 391



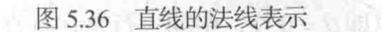


图 5.36　直线的法线表示

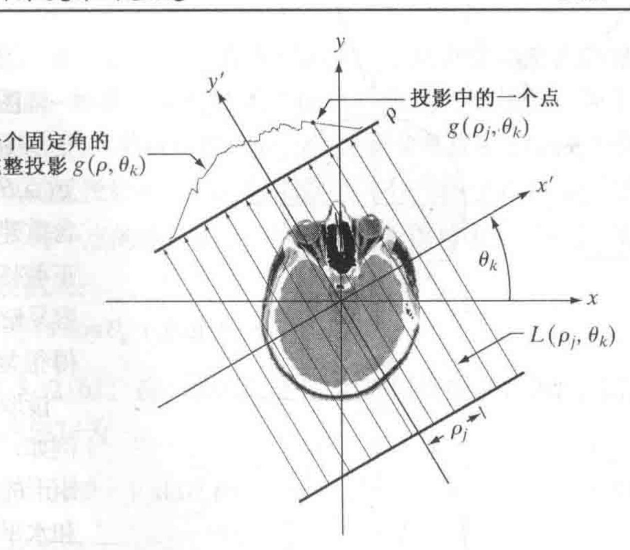


图 5.37　平行射线束的几何描述

在离散情况下，式(5.11-3)变成

$$g(\rho,\theta)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\delta(x\cos\theta+y\sin\theta-\rho) \tag{5.11-4}$$

式中，x, y, ρ, θ 现在是离散变量。如果固定 θ 而令 ρ 变化，我们可以看到，式(5.11-4)简化为沿着由这两个参数的特定值定义的线对 $f(x, y)$ 的像素求和。(固定 θ 值的情况下)通过增加覆盖图像(θ 值固定)所要求的 ρ 的所有值产生一个投影。改变 θ 并重复前述过程则产生另一个投影，以此类推。这正好就是 5.11.1 节中生成投影的方法。

例 5.17　使用雷登变换得到圆形区域的投影。

在论述之前，我们首先说明如何使用雷登变换得到图5.38(a)中圆形物体投影的解析表达式：

$$f(x,y)=\begin{cases}A, & x^2+y^2\leqslant r^2\\ 0, & \text{其他}\end{cases}$$

式中，A 是常数，r 是物体的半径。我们假设圆心位于 xy 平面的原点。因为物体是圆对称的，其投影对所有角度都是一样的，所以我们要做的全部事情就是得到 $\theta=0°$ 的投影。然后，式(5.11-3)变成

$$\begin{aligned}g(\rho,\theta)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(x-\rho)\mathrm{d}x\,\mathrm{d}y\\&=\int_{-\infty}^{\infty}f(\rho,y)\mathrm{d}y\end{aligned}$$

其中，第二行来自式(4.2-10)。如之前注释的那样，这是一个线积分(在这种情况下沿着线 $L(\rho,0)$ 积分)。还要注意的是，当 $|\rho|>r$ 时，$g(\rho,\theta)=0$。当 $|\rho|\leqslant r$ 时，积分从 $y=-\sqrt{r^2-\rho^2}$ 到 $y=\sqrt{r^2-\rho^2}$ 求值。因此，

$$g(\rho,\theta)=\int_{-\sqrt{r^2-\rho^2}}^{\sqrt{r^2-\rho^2}}f(\rho,y)\mathrm{d}y=\int_{-\sqrt{r^2-\rho^2}}^{\sqrt{r^2-\rho^2}}A\mathrm{d}y$$

执行积分得到

$$g(\rho,\theta)=g(\rho)=\begin{cases}2A\sqrt{r^2-\rho^2}, & |\rho|\leqslant r\\ 0, & \text{其他}\end{cases}$$

其中，我们利用了 $|\rho|>r$ 时 $g(\rho,\theta)=0$ 这样一个事实。图 5.38(b) 显示了这一结果，该结果与图 5.32 和图 5.33 中所说明的投影一致。注意，$g(\rho,\theta)=g(\rho)$，即 g 与 θ 无关，因为物体是关于原点对称的。

392～393

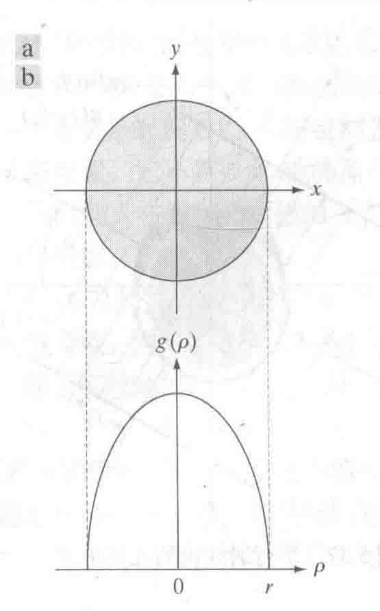
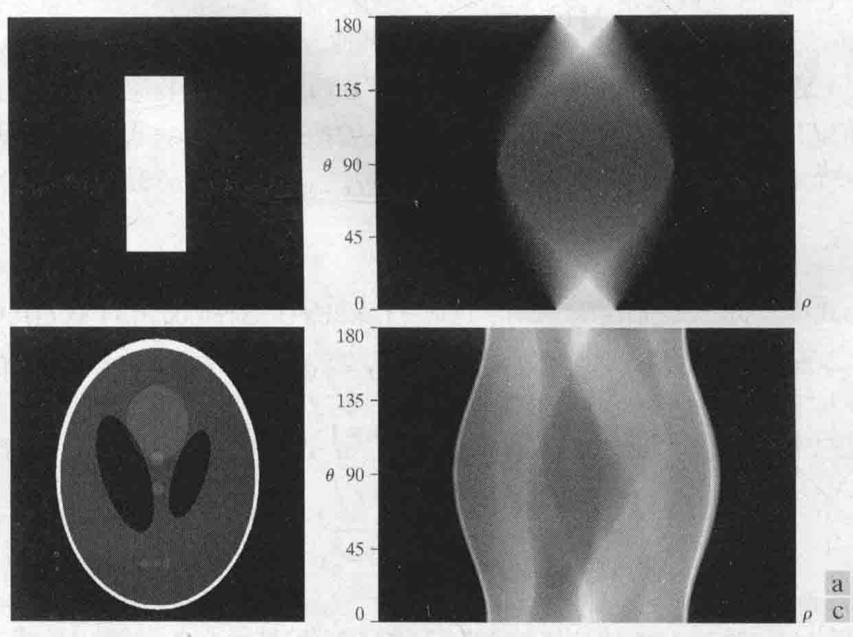


图 5.38　(a)一个圆盘；(b)该圆盘解析推导的雷登变换的曲线。这里我们可以画出该变换，因为它仅取决于一个变量。当 g 取决于 ρ 和 θ 时，雷登变换变成一幅图像，其坐标轴是 ρ 和 θ，而一个像素的亮度与该像素位置处的 g 值成正比

当雷登变换 $g(\rho,\theta)$ 以 ρ 和 θ 作为直线坐标显示为一幅图像时，结果称之为正弦图(sinogram)，概念上类似于所显示的傅里叶谱［但与傅里叶变换不同的是，$g(\rho,\theta)$ 总是一个实函数］。类似于傅里叶变换，正弦图包含重建 $f(x,y)$ 所需的数据。如同显示的傅里叶谱一样，正弦图对于简单区域很容易解释，但作为一个变得很复杂的投影区域去“读取”会变得很困难。例如，图 5.39(b)是左侧矩形的正弦图，其垂直轴和水平轴分别对应于 ρ 和 θ。因此，底部一行是矩形在水平方向上的投影(即 $\theta = 0°$)，中间一行是在垂直方向上的投影(即 $\theta = 90°$)。底部一行的非零部分比中间一行的非零部分小这一事实告诉我们，物体在水平方向较窄。正弦图在两个方向上关于图像中心对称的事实告诉我们，处理的物体是对称的，并且平行于 x 轴和 y 轴。最后，正弦图是平滑的，这表明物体具有均匀的灰度。除了这些类型的一般观察外，关于正弦图我们说不出更多的特点。

> 为生成有着相同行数的阵列，正弦图中 ρ 轴的最小维对应于投影期间遇到的最大维。例如，使用 1°增量得到的大小为 $M \times M$ 的方形的正弦图的最小尺寸为 $180 \times Q$，其中 Q 大于 $\sqrt{2}M$ 的最小整数。

图5.39(c)显示了一幅 Shepp-Logan 幻影图像，该图像是广泛使用的合成图像，设计用于模拟脑部主要区域的吸收情况，包括几个小肿瘤。如图5.39(d)所示，解释这幅图像的正弦图更加困难。我们仍然可以推断某些对称的特点，但这是我们能够说明的全部内容。正弦图的视觉分析仅限于实际应用，但有时对于开发算法是有帮助的。



393 ~ 394

图 5.39　两幅图像和它们的正弦图(雷登变换)。正弦图的每一行是在垂直轴上沿相应角度的投影。图像(c)称为 Shepp-Logan 幻影。在其原始形式中，幻影的灰度非常低。为便于观察，这里显示的是增强后的图像

CT 的关键目的是从投影得到物体的三维表示。如 5.11.1 节直观介绍的那样，其方法是反投影每个投影，然后对反投影求和来产生一幅图像（切片）。堆积所有的结果图像后，再现三维物体。为了得到来自雷登变换的反投影图像的形式化表达，让我们从单点开始，$g(\rho_j,\theta_k)$ 是全部投影 $g(\rho,\theta_k)$ 的一个固定旋转值 θ_k 的投影（见图 5.37）。由这个单点的反投影形成的一幅图像的部分与将线 $L(\rho_j,\theta_k)$ 复制到图像上相比，并不能提供更多的信息，其中线上每一点的值是 $g(\rho_j,\theta_k)$。对投影信号中的所有 ρ_j 值重复这一过程（保持 θ 值固定为 θ_k），可得到如下表达式：

$$f_{\theta_k}(x,y)=g(\rho,\theta_k)=g(x\cos\theta_k+y\sin\theta_k,\theta_k)$$

对固定角度 θ_k 投影进行反投影所得到的图像见图 5.32(b)。这一公式对于任意 θ_k 值均成立，因此我们一般可以把由角度 θ 处得到的单个反投影形成的图像写为

$$f_\theta(x,y)=g(x\cos\theta+y\sin\theta,\theta) \tag{5.11-5}$$

通过对所有的反投影图像积分，我们得到最终图像：

$$f(x,y)=\int_0^\pi f_\theta(x,y)\mathrm{d}\theta \tag{5.11-6}$$

在离散情况下，积分变成对所有反投影图像求和：

$$f(x,y)=\sum_{\theta=0}^{\pi} f_\theta(x,y) \tag{5.11-7}$$

式中，x, y 和 θ 现在是离散值。回顾 5.11.1 节中的讨论，在 0°和 180°处的投影互为镜像图像，因此求和执行最后的角度增量在 180°之前。例如，如果使用 0.5°的增量，则求和操作是从 0°到 179.5°以半度为增量。用刚刚讨论的方法形成的反投影图像有时称为*层图*（laminogram）。我们可以隐含地将层图理解为仅是对产生投影的图像的一个近似，下面的例子将清楚地解释这一事实。

例 5.18　从正弦图得到反投影图像。

式(5.11-7)用于从式(5.11-4)得到的投影产生图 5.32 到图 5.34 的反投影图像。类似地，这些公式用于产生图 5.40(a)和图 5.40(b)，它们分别显示了对应于图 5.39(b)和图 5.39(d)中正弦图的反投影图像。如之前的图像那样，我们注意到出现了很严重的模糊现象，因此很明显，直接运用式(5.11-4)和式(5.11-7)将得到不可接受的结果。然而，如 5.11.5 节中介绍的那样，采用重新表达反投影的方法，有可能改善重建的效果。

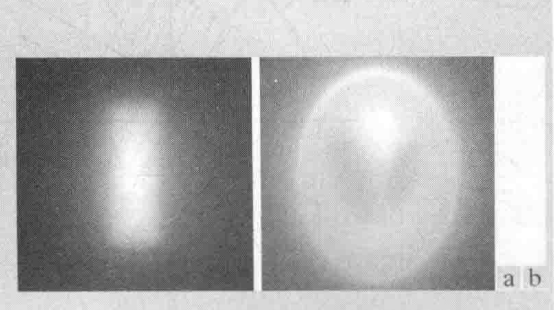


图 5.40　图 5.39 中的正弦图的反投影

5.11.4　傅里叶切片定理

本节推导与投影的一维傅里叶变换和得到投影区域的二维傅里叶变换相关的基本结果。这一关系是能处理刚才讨论的模糊问题的重建方法的基础。

关于 ρ 投影的一维傅里叶变换为

$$G(\omega,\theta)=\int_{-\infty}^{\infty} g(\rho,\theta)\mathrm{e}^{-\mathrm{j}2\pi\omega\rho}\mathrm{d}\rho \tag{5.11-8}$$

其中，如式(4.2-16)所示，ω 是频率变量，很容易理解该表达式是给定 θ 值的。用式(5.11-3)代替

394 ~ 396

$g(\rho,\theta)$，得到表达式

$$G(\omega,\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(x\cos\theta+y\sin\theta-\rho)\mathrm{e}^{-\mathrm{j}2\pi\omega\rho}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\rho$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\left[\int_{-\infty}^{\infty}\delta(x\cos\theta+y\sin\theta-\rho)\mathrm{e}^{-\mathrm{j}2\pi\omega\rho}\mathrm{d}\rho\right]\mathrm{d}x\,\mathrm{d}y \tag{5.11-9}$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi\omega(x\cos\theta+y\sin\theta)}\,\mathrm{d}x\,\mathrm{d}y$$

其中，最后一步遵循本节早些时候提到过的冲激特性。令 $u=\omega\cos\theta$ 和 $v=\omega\sin\theta$，式(5.11-9)变为

$$G(\omega,\theta)\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\mathrm{e}^{-\mathrm{j}2\pi(ux+vy)}\,\mathrm{d}x\,\mathrm{d}y\right]_{u=\omega\cos\theta;\,v=\omega\sin\theta} \tag{5.11-10}$$

我们可以看出该表达式与指定 u 值和 v 值时 $f(x,y)$［见式(4.5-7)］的二维傅里叶变换等价，即

$$G(\omega,\theta)=\left[F(u,v)\right]_{u=\omega\cos\theta;\,v=\omega\sin\theta}=F(\omega\cos\theta,\omega\sin\theta) \tag{5.11-11}$$

其中，$F(u,v)$ 通常表示 $f(x,y)$ 的二维傅里叶变换。

式(5.11-11)就是众所周知的傅里叶切片定理(或投影切片定理)。它说明一个投影的傅里叶变换是得到投影区域的二维傅里叶变换的一个切片。该术语的来源可用图5.41来解释。如该图所示，任意一个投影的一维傅里叶变换可沿一个角度提取一条线的 $F(u,v)$值得到，该角度是产生投影所用的角度。原理上，我们可以简单地使用 $F(u,v)$的傅里叶反变换得到 $f(x,y)$[①]，但这是高代价的计算，因为它涉及二维变换的反变换。下一节中讨论的方法将有效得多。

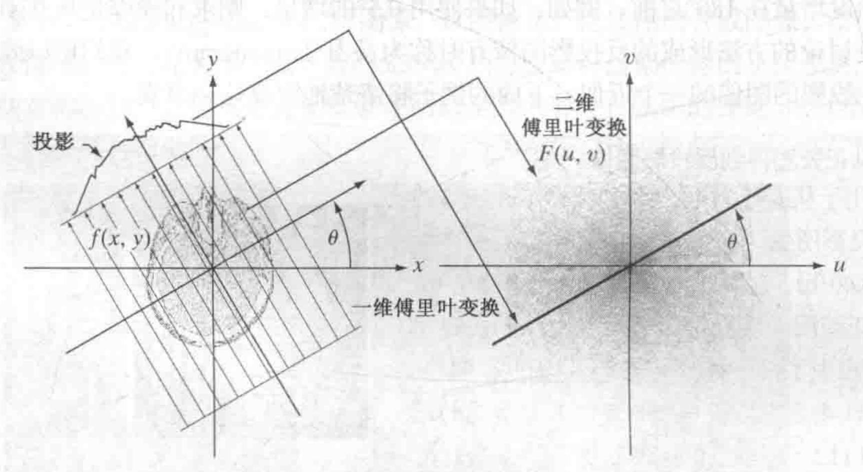


图5.41 傅里叶切片定理的说明。投影的一维傅里叶变换是得到投影的区域的二维傅里叶变换的一个切片。注意角度θ的对应

5.11.5 使用平行射线束滤波反投影的重建

如我们在5.11.1节和例5.18中看到的那样，直接得到反投影会生成不可接受的模糊结果。所幸的是，解决该问题有一种简单的方法，即在计算反投影前对投影做简单的滤波。从式(4.5-8)可知，$F(u,v)$的二维反傅里叶变换是

396
~
397

① 记住，即使使用了傅里叶反变换，恢复的图像中仍会存在模糊，因为该结果等同于使用前一节讨论的方法得到的结果。

$$f(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F(u,v)\mathrm{e}^{\mathrm{j}2\pi(ux+vy)}\mathrm{d}u\ \mathrm{d}v \tag{5.11-12}$$

> $\mathrm{d}u\mathrm{d}v=\omega\mathrm{d}\omega\mathrm{d}\theta$ 这一关系源自基本积分学，在基本积分学中，雅可比矩阵是各变量变化的基础。

如式(5.11-10)和式(5.11-11)所示，若令 $u=\omega\cos\theta$ 和 $v=\omega\sin\theta$，则微分变为 $\mathrm{d}u\mathrm{d}v=\omega\mathrm{d}\omega\mathrm{d}\theta$，我们可以把式(5.11-12)表达为极坐标形式

$$f(x,y)=\int_{0}^{2\pi}\int_{0}^{\infty}F(\omega\cos\theta,\omega\sin\theta)\mathrm{e}^{\mathrm{j}2\pi\omega(x\cos\theta+y\sin\theta)}\omega\ \mathrm{d}\omega\ \mathrm{d}\theta \tag{5.11-13}$$

然后，利用傅里叶切片定理有

$$f(x,y)=\int_{0}^{2\pi}\int_{0}^{\infty}G(\omega,\theta)\mathrm{e}^{\mathrm{j}2\pi\omega(x\cos\theta+y\sin\theta)}\omega\ \mathrm{d}\omega\ \mathrm{d}\theta \tag{5.11-14}$$

把积分分离为两个表达式，一个表达式的θ 取值范围为 0°到 180°，另一表达式的θ 取值范围为 180°到 360°，并利用$G(\omega,\theta+180°)=G(-\omega,\theta)$这一事实(见习题 5.32)，可以把式(5.11-14)表达为

$$f(x,y)=\int_{0}^{\pi}\int_{-\infty}^{\infty}|\omega|G(\omega,\theta)\mathrm{e}^{\mathrm{j}2\pi\omega(x\cos\theta+y\sin\theta)}\ \mathrm{d}\omega\ \mathrm{d}\theta \tag{5.11-15}$$

考虑关于ω 的积分项，由式(5.11-1)可以看出，如ρ 那样，$x\cos\theta+y\sin\theta$ 是常数。因此，式(5.11-15)可写为

$$f(x,y)=\int_{0}^{\pi}\left[\int_{-\infty}^{\infty}|\omega|G(\omega,\theta)\mathrm{e}^{\mathrm{j}2\pi\omega\rho}\mathrm{d}\omega\right]_{\rho=x\cos\theta+y\sin\theta}\mathrm{d}\theta \tag{5.11-16}$$

方括号内的表达式是一维傅里叶反变换形式［见式(4.2-17)］，只是附加了一个$|\omega|$项，基于 4.7 节的讨论，我们可以看出它是一个一维滤波函数。显然，$|\omega|$是一个斜坡滤波器［见图5.42(a)］[①]。这个函数不可积，因为其幅度在两个方向上都扩展到了 $+\infty$，所以其傅里叶反变换无定义。理论上，可用某种方法来处理该问题，例如使用所谓的广义δ 函数。实践中，该方法是对斜坡加窗，使它在定义的频率范围之外为零，也就是加窗限制带宽的斜坡滤波器。

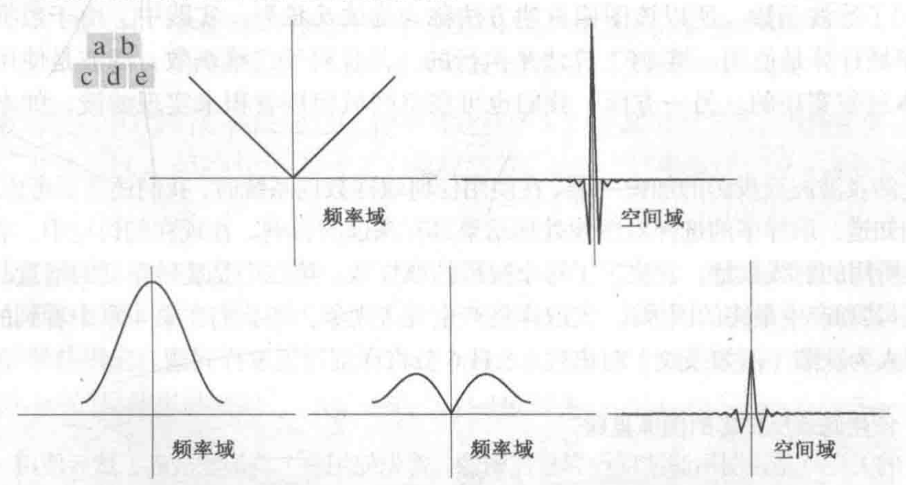


图 5.42　(a)使用一个盒状滤波器限制带宽后的滤波器$|\omega|$的频率域剖面线；(b)空间域表示；(c) 汉明窗函数；(d)由图(a)和图(c)的乘积形成的加窗后的斜坡滤波器；(e)乘积的空间域表示(注意振铃的降低)

限制一个函数的带宽的最简方法是，在频率域中使用一个盒子。然而，正如我们在图 4.4 中看到的那样，一个盒子有不希望的振铃特性，因此用一个平滑窗来代替。图 5.42(a)显示了使用一个

397~399

① 斜坡滤波器通常称为 Ram-Lak 滤波器，因为该滤波器是由 Ramachandran and Lakshminarayanan[1971]首次提出的。

盒状窗限制带宽后的斜坡滤波器，图 5.42(b)显示了计算该滤波器的傅里叶反变换得到的空间域表示。如预期的那样，加窗滤波的结果在空间域中呈现出了明显的振铃。我们由第 4 章可知，频率域的滤波等于空间域的卷积，因此，使用存在振铃的函数进行空间滤波也将产生被振铃污染的结果。使用平滑函数加窗有助于这一情况。常用于实现一维 FFT 的 M 点离散窗函数由下式给出：

$$h(\omega)=\begin{cases} c+(c-1)\cos\dfrac{2\pi\omega}{M-1}, & 0\leqslant\omega\leqslant(M-1) \\ 0, & \text{其他} \end{cases} \tag{5.11-17}$$

当 $c=0.54$ 时，该函数称为汉明窗［根据 Richard Hamming(理查德·汉明)命名］；当 $c=0.5$ 时，该函数称为韩窗［根据 Julius von Hann(朱利叶·冯·韩)命名］[①]。汉明窗和韩窗间的主要区别是，后者末尾的一些点为零。通常，两者间的差别在图像处理应用中是觉察不到的。

图 5.42(c)显示了汉明窗的曲线，图 5.42(d)显示了汉明窗和图 5.42(a)中带限斜坡滤波器的积。图 5.42(e)显示了该积在空间域的表示，这种表示通常是通过计算其反 FFT 得到的。很明显，该图与图 5.42(b)比较，加窗斜坡中的振铃现象减少了［图5.42(b)和图5.42(e)中的峰值比分别为 2.5 和 3.4］。另一方面，因为图 5.42(e)中的中心瓣比图 5.42(b)中的中心瓣稍宽，我们可以预期，由于使用汉明窗的反投影有较小的振铃，但稍微多一点模糊。如例 5.19 所示，情况的确如此。

回顾式(5.11-8)，$G(\omega,\theta)$ 是 $g(\rho,\theta)$ 的一维傅里叶变换，这是在一个固定角度 θ 得到的单一投影。式(5.11-16)说明完整的反投影图像 $f(x,y)$ 是由如下步骤得到的：

1. 计算每个投影的一维傅里叶变换。
2. 如上面说明的那样，用滤波函数 $|\omega|$ 乘以每个傅里叶变换，即乘以一个合适的窗(如汉明窗)。
3. 得到每个滤波后的变换的一维反傅里叶变换。
4. 对步骤 3 得到的所有一维反变换积分(求和)。

因为使用了滤波函数，所以该图像重建方法称为*滤波反投影*。实践中，由于数据是离散的，因此所有频率域计算是使用一维 FFT 算法来执行的，并且对于二维函数，滤波是使用第 4 章中说明的相同基本过程实现的。另一方面，我们也可在空间域使用卷积来实现滤波，如本节稍后说明的那样。

前述讨论涉及滤波反投影的加窗问题。在使用任何取样数据系统时，我们还需要考虑取样率问题。由第 4 章我们知道，取样率的选择对图像处理结果具有深远的影响。在现在的讨论中，有两个取样考虑。第一个是所用的射线数量，它决定了每个投影的取样数。第二个是旋转角度的增量，它决定了重建图像的数量(其和产生最终的图像)。欠取样将产生混淆现象，如我们在第 4 章中看到的那样，它在图像本身出现人为缺陷（例如条纹）时出现。5.11.6 节将详细讨论取样问题。

例 5.19 使用滤波反投影的图像重建。

这个例子的关注点是说明用滤波反投影重建图像，首先使用一个斜坡滤波器，然后使用一个被汉明窗改进的斜坡滤波器，并将这些滤波反投影与图5.40中的原始反投影结果相比较。为了只关注滤波的差别，在这个例子中，结果是使用 0.5°的旋转增量产生的，它是我们在图5.40中用过的增量。在两种情况下，射线间的间隔为一个像素。两个例子中的图像尺寸均为 600×600 像素，因此，对角线的长度是 $\sqrt{2}\times600\approx849$。当旋转角度是 45°和 135°时，所用的 849 条射线覆盖了整个区域。

399 ~ 400

① 有时，为与汉明窗对比，韩窗也称为汉宁窗。但汉宁窗这一术语是不正确的，是引起混乱的根源。

图5.43(a)显示了使用斜坡滤波器的矩形重建。这一结果中的最清晰特性是没有视觉上可见的模糊。然而，如预期的那样，它存在振铃现象，我们能见到暗淡的线条，特别是在矩形四个角的周围。这些线条在图5.43(c)的放大部分中看得更清楚。在斜坡滤波器上使用汉明窗对振铃问题有相当大的帮助，如图 5.43(b)和图 5.43(d)所示，但要付出稍微有点模糊的代价。对图 5.43(a)的改进很明显(与未加窗的斜坡滤波器相比)。由于幻影图像没有像矩形一样急剧且突出的过渡，因此即使是使用未加窗的斜坡滤波器，振铃现象也是感觉不到的，正如图 5.44(a)所示的那样。使用汉明窗会得到稍微平滑一些的图像，如图 5.44(b)所示。与图 5.40(b)相比，两个结果都有较大的改进，这再次证明了滤波反投影方法所固有的优点。

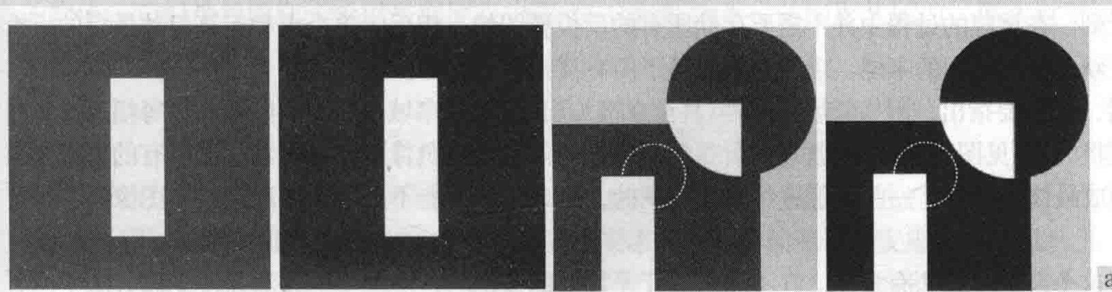


图 5.43　(a)使用斜坡滤波器的矩形滤波反投影；(b)用汉明窗斜坡滤波器的矩形滤波反投影。第二行显示了第一行中图像放大后的细节。请与图 5.40(a)比较

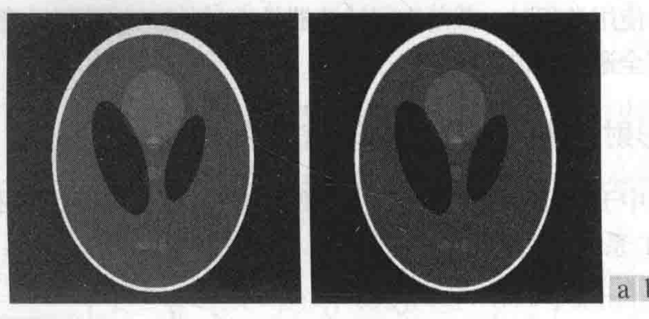


图 5.44　(a)使用斜坡滤波器的头部幻影的滤波反投影；(b)使汉明窗斜坡滤波器的头部幻影的滤波反投影。请与图 5.40(b)比较

在 CT 的多数应用中(特别是在医学上)，像振铃这样的人为缺陷影响严惩，因此须使它们最小化。调整滤波算法，并像 5.11.2 节说明的那样使用大量检测器，是在设计中有助于减少这些影响应考虑的事项。

前述讨论是以通过一个 FFT 实现得到滤波反投影为基础的。然而，由第 4 章中的卷积理论可知，使用空间卷积也可得到相同的结果。特别地，式(5.11-16)中方括号内的项是两个频率域函数乘积的反傅里叶变换，根据卷积定理，我们知道它等于这两个函数的空间表示(反傅里叶变换)的卷积。换句话说，令 $s(\rho)$ 表示 $|\omega|$ 的反傅里叶变换[①]，我们将式(5.11-16)写为

$$\begin{aligned} f(x,y) &= \int_0^{\pi}\left[\int_{-\infty}^{\infty}|\omega|G(\omega,\theta)\mathrm{e}^{\mathrm{j}2\pi\omega\rho}\mathrm{d}\omega\right]_{\rho=x\cos\theta+y\sin\theta}\mathrm{d}\theta = \int_0^{\pi}\left[s(\rho)\bigstar g(\rho,\theta)\right]_{\rho=x\cos\theta+y\sin\theta}\mathrm{d}\theta \\ &= \int_0^{\pi}\left[\int_{-\infty}^{\infty}g(\rho,\theta)s(x\cos\theta+y\sin\theta-\rho)\mathrm{d}\rho\right]\mathrm{d}\theta \end{aligned} \tag{5.11-18}$$

① 如果使用了一个窗函数，如汉明窗，则反傅里叶变换对加窗的斜坡函数执行。同样，我们再次忽略了前面提及的关于连续反傅里叶变换的存在性问题，因为所有的实现均是使用有限长的离散量执行的。

400 ~ 402

其中，如第 4 章中那样，★表示卷积。第二行的原因见前一段的解释。第三行满足式(4.2-20)给出的卷积的实际定义。

式(5.11-18)的后两行描述了相同的事情：角度 θ 处的各个反投影可以用相应的投影 $g(\rho,\theta)$ 和斜坡滤波器 $s(\rho)$ 的反傅里叶变换的卷积操作得到。与以前一样，整个反投影图像可通过对各个反投影图像积分(求和)得到。除了计算上的舍入误差外，使用卷积的结果与使用 FFT 的结果相同。在实际的 CT 实现中，已证明卷积通常在计算上更有效，因此，多数现代 CT 系统使用这种方法。傅里叶变换在理论表示和算法开发中确实扮演了重要角色(例如，MATLAB 中的 CT 图像处理就是基于 FFT 的)。另外，我们注意到，在重建的过程中并不需要存储所有的反投影图像。相反，单个求和运算仅被最后的反投影图像更新。在该过程的末尾，求和运算将等于所有反投影的总和。

最后，我们要指出，因为斜坡滤波器(甚至在被加窗时)在频率域的直流项归零，故每幅反投影图像的均值将为零(见图 4.30)。这就意味着每幅反投影图像都将有负像素和正像素。当所有的反投影图像相加形成最终图像时，一些负位置可能变成正的，而平均值可能不是零，但通常最终图像仍将有负像素。

处理这个问题的方法有多种。在不了解关于平均值的知识时，最简单的方法是接受该方法中固有负值的这一事实，并使用式(2.6-10)和式(2.6-11)中描述的过程来标定这一结果。这也是本节采用的方法。具备典型平均值的知识时，可将该值加到频率域的滤波器上，以便抵消斜坡并防止直流项为零[见图 4.31(c)]。在空间域中使用卷积时，截断空间滤波器长度(斜坡的反傅里叶变换)的真正效果都可防止其有零均值，这样就完全避免了迫零问题。

5.11.6 使用扇形射线束滤波反投影的重建

迄今为止的讨论都集中于平行射线束。因为它简单而直观，这是传统上介绍计算机断层成像的成像几何原理。然而，现代 CT 系统却使用扇形射线束的几何原理(见图5.35)，因此本节剩下部分将讨论这一主题。

图 5.45 显示了一个基本的扇形射线束成像几何，其中检测器安放在一个圆弧上，并假设源的角度增量相等。令 $p(\alpha,\beta)$ 表示一个扇形射线束投影，其中 α 是相对于中心射线测量特定检测器的角度，β 是相对于 y 轴测量的源的角度位移，如图所示。在图 5.45 中，我们还注意到，扇形射线束中的一条射线可用一条线 $L(\rho,\theta)$ 来表示，在标准形式中，它是我们前一节讨论平行射线束成像几何中用以描述一条射线的方法。这就允许我们用平行射线束的结果作为推导扇形射线几何公式的起点。我们继续以卷积为基础推导扇形射线滤波反投影的方法。[1]

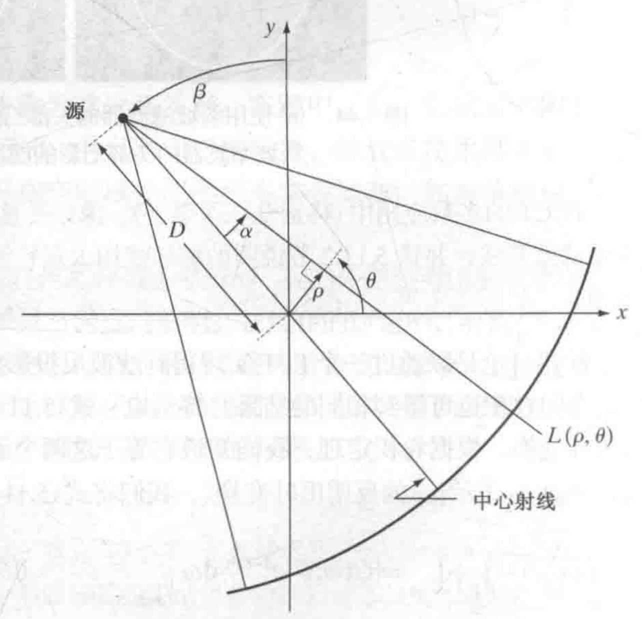


图 5.45 基本的扇形射线几何原理。通过源的中心和原点(这里假设为源的旋转中心)的线称为中心射线

[1] 为平行射线束几何原理推导的傅里叶切片定理并不能直接用于扇形射线束。但是，式(5.11-19)和式(5.11-20)奠定了将扇形射线束几何转换为平行射线束几何的基础，从而允许我们使用前一节中推导的滤波后的平行反投影方法，这时切片定理是适用的。本节的末尾将详细探讨这一点。

402~404

我们以图 5.45 中的表示开始，线 $L(\rho, \theta)$ 的参数与扇形射线参数的关系是

$$\theta = \beta + \alpha \tag{5.11-19}$$

和

$$\rho = D \sin \alpha \tag{5.11-20}$$

式中，D 是从源的中心到 xy 平面的原点的距离。

平行射线成像几何的卷积反投影公式已由式(5.11-18)给出。不失一般性，假设我们集中注意这样的物体，它们被包围在一个以平面原点为中心、以 T 为半径的圆形区域内。那么对于 $|\rho| > T$，有 $g(\rho,\theta) = 0$，且式(5.11-18)变为

$$f(x,y) = \frac{1}{2}\int_0^{2\pi}\int_{-T}^{T} g(\rho,\theta)s(x\cos\theta + y\sin\theta - \rho)\mathrm{d}\rho\,\mathrm{d}\theta \tag{5.11-21}$$

其中，我们利用了 5.11.1 节说明的事实，即相隔 180°的投影互为镜像。在这种方法中，式(5.11-21)中的外积分限扫过整个圆，正如扇形射线束配置所要求的那样，检测器被安排在一个圆上。

我们感兴趣的是关于 α 和 β 的积分。为这样做，我们将它改为极坐标 (r,φ)，即令 $x = r\cos\varphi$ 和 $y = r\sin\varphi$，由此，它变为下式：

$$x\cos\theta + y\sin\theta = r\cos\varphi\cos\theta + r\sin\varphi\sin\theta = r\cos(\theta - \varphi) \tag{5.11-22}$$

利用这一结果，我们可把式(5.11-21)表示为

$$f(x,y) = \frac{1}{2}\int_0^{2\pi}\int_{-T}^{T} g(\rho,\theta)s[r\cos(\theta-\varphi) - \rho]\mathrm{d}\rho\,\mathrm{d}\theta$$

这个表达式的内容并不比平行射线束重建写成极坐标的公式更多，但积分还是关于 ρ 和 θ 的。为了对 α 和 β 积分，要求用式(5.11-19)和式(5.11-20)进行坐标变换：

$$f(r,\varphi) = \frac{1}{2}\int_{-\alpha}^{2\pi-\alpha}\int_{\arcsin(-T/D)}^{\arcsin(T/D)} g(D\sin\alpha, \alpha+\beta)s\left[r\cos(\beta+\alpha-\varphi) - D\sin\alpha\right]D\cos\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{5.11-23}$$

式中，我们使用了 $\mathrm{d}\rho\,\mathrm{d}\theta = D\cos\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta$ ［见式(5.11-13)的解释］。

这个公式可以进一步简化。首先，注意积分限从 $-\alpha$ 到 $2\pi - \alpha$，对 β 来说，扫过了全部 360 度。因为 β 的所有函数是周期为 2π的周期函数，外积分限可分别用 0 和 2π 来代替。项 $\arcsin(T/D)$ 对应于 $|\rho| > T$ 有一个最大值 α_m，超过该值则有 $g = 0$(见图 5.46)，因此，我们可以把内积分的积分限分别用 $-\alpha_m$ 和 α_m 来代替。最后考虑图5.45 中的线 $L(\rho,\theta)$。沿这条线的扇形射线和，必须等于沿同一条线的平行射线和(射线和是沿一条线的所有值的和，因此对于一条给定的射线，结果必须相同，不管在什么坐标系下)。对于 (α,β) 和 (ρ,θ) 对应值的任何射线和，这都是正确的。这样，令 $p(\alpha,\beta)$ 表示一个扇形射线投影，它满足 $p(\alpha,\beta) = g(\rho,\theta)$，且由式(5.11-19)和式(5.11-20)有 $p(\alpha,\beta) = g(D\sin\alpha, \alpha+\beta)$。合并这两个等式为式(5.11-23)，我们有

$$f(r,\varphi) = \frac{1}{2}\int_0^{2\pi}\int_{-\alpha_m}^{\alpha_m} p(\alpha,\beta)s\left[r\cos(\beta+\alpha-\varphi) - D\sin\alpha\right]D\cos\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{5.11-24}$$

这是基于滤波反投影的基本扇形重建公式。

式(5.11-24)可进一步变换为更熟悉的卷积形式。参考图 5.47，它可表示为(见习题 5.33)

$$r\cos(\beta+\alpha-\varphi) - D\sin\alpha = R\sin(\alpha' - \alpha) \tag{5.11-25}$$

式中，R 是在扇形射线中源到任何一点的距离，α' 是该条射线与中心射线的夹角。注意，R 和 α' 由 r，φ 和 β 的值决定。把式(5.11-25)代入式(5.11-24)得到

404 ~ 406

$$f(r,\varphi)=\frac{1}{2}\int_{0}^{2\pi}\int_{-\alpha_m}^{\alpha_m}p(\alpha,\beta)s\left[R\sin(\alpha'-\alpha)\right]D\cos\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta \tag{5.11-26}$$

可以证明(见习题 5.34)

$$s(R\sin\alpha)=\left(\frac{\alpha}{R\sin\alpha}\right)^{2}s(\alpha) \tag{5.11-27}$$

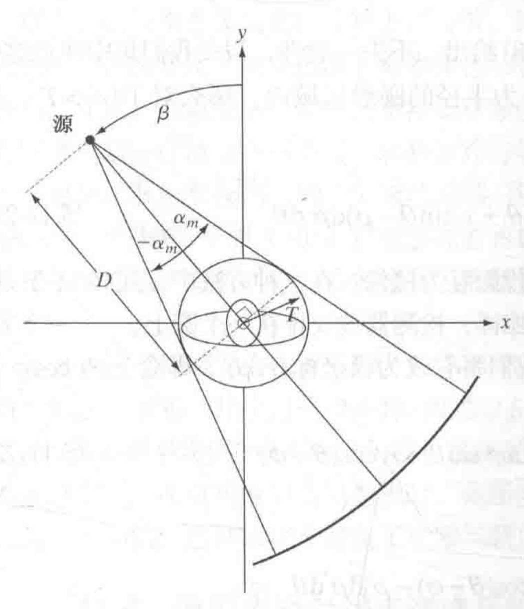



图 5.46 包围感兴趣区域所需的最大α值

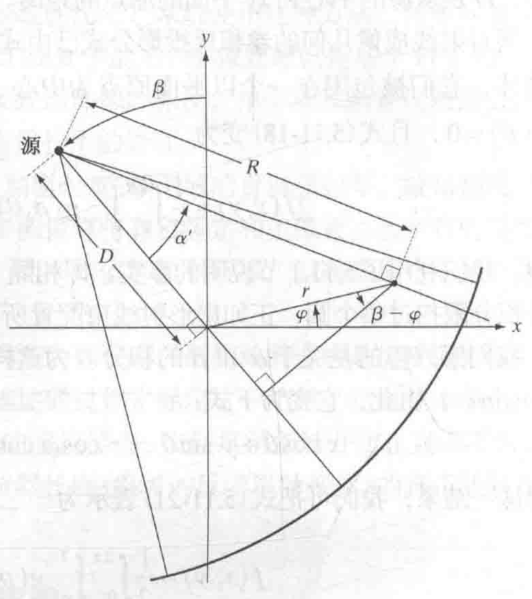


图 5.47 扇形射线上任意一点的极坐标表示

使用该表达式，可以把式(5.11-26)写为

$$f(r,\varphi)=\int_{0}^{2\pi}\frac{1}{R^{2}}\left[\int_{-\alpha_m}^{\alpha_m}q(\alpha,\beta)h(\alpha'-\alpha)\mathrm{d}\alpha\right]\mathrm{d}\beta \tag{5.11-28}$$

式中，

$$h(\alpha)=\frac{1}{2}\left(\frac{\alpha}{\sin\alpha}\right)^{2}s(\alpha) \tag{5.11-29}$$

和

$$q(\alpha,\beta)=p(\alpha,\beta)D\cos\alpha \tag{5.11-30}$$

我们可以看出式(5.11-28)中方括号内的积分是一个卷积表达式，这样，式(5.11-24)中的图像重建公式就可以如同用函数$q(\alpha,\beta)$和$h(\alpha)$的卷积那样来实现。与平行投影重建的公式不同，基于扇形射线束投影重建包含有 $1/R^2$ 项，它是与源的距离成反比的加权系数。实现式(5.11-28)的计算细节超出了当前讨论的范围(关于该主题的详细论述，请参阅 Kak and Slaney[2001])。

直接代替式(5.11-28)的实现，特别是在软件模拟中的一种常用方法是：(1)使用式(5.11-19)和式(5.11-20)把扇形射线几何转换为平行射线几何，(2)使用 5.11.5 节中开发的平行射线重建方法。我们将用一个如何这样做的例子来结束本节的讨论。如早些时候指出的那样，以角度β获取的一个扇形射线投影p，有一个相应的以角度θ获取的平行射线投影g，因此有

$$p(\alpha,\beta)=g(\rho,\theta)=g(D\sin\alpha,\alpha+\beta) \tag{5.11-31}$$

其中，最后一项来自式(5.11-19)和式(5.11-20)。

406 ~ 407

令$\Delta\beta$表示连续扇形投影的角度增量，并令$\Delta\alpha$是射线之间的角度增量，它决定每个投影中的样本数量。利用约束

$$\Delta\beta = \Delta\alpha = \gamma \tag{5.11-32}$$

然后，利用 $\beta = m\gamma$ 和 $\alpha = n\gamma$，其中 m 和 n 是整数，可以把式(5.11-31)写成

$$p(n\gamma, m\gamma) = g\left[D\sin n\gamma, (m+n)\gamma\right] \tag{5.11-33}$$

该式指出第 m 个射线投影中的第 n 条射线，等于第$(m+n)$个平行投影中的第 n 条射线。式(5.11-33)右边的 $D\sin n\gamma$ 项意味着从扇形射线投影转换为平行射线投影是不均匀取样，如果取样间隔 $\Delta\alpha$ 和 $\Delta\beta$ 太粗，则可能导致模糊、振铃和混淆问题，如下例所述。

例 5.20　使用滤波扇形反投影的图像重建。

图 5.48(a)显示了：(1)使用 $\Delta\alpha = \Delta\beta = 1°$ 产生矩形图像的扇形投影，(2)使用式(5.11-33)将每个扇形射线转换为相应的平行射线，(3)把5.11.5节推出的滤波反投影方法用于平行射线。图5.48(b)到(d)分别显示了使用0.5°、0.25°和0.125°增量的结果。所有情况下都使用了汉明窗。这一角度增量的变化用于说明欠取样的效应。

图5.48(a)中的结果清楚地指出，1°的增量太粗，模糊和振铃现象十分明显。图5.48(b)中的结果在感觉上很有趣，它与图5.43(b)相比很差，而图5.43(b)也是使用0.5°的角度增量产生的。事实上，如图5.48(c)所示，即便使用0.25°的角度增量重建也不比图5.43(b)好。如图5.48(d)所示，为使两个结果可以比较，我们必须使用约为0.125°的角度增量。这个角度增量可产生180×(1/0.25)=720个样本，这接近于例5.19中的平行投影所用的849条射线。这样，可以预料这些结果在外观上接近于使用 $\Delta\alpha = 0.125°$ 时的结果。

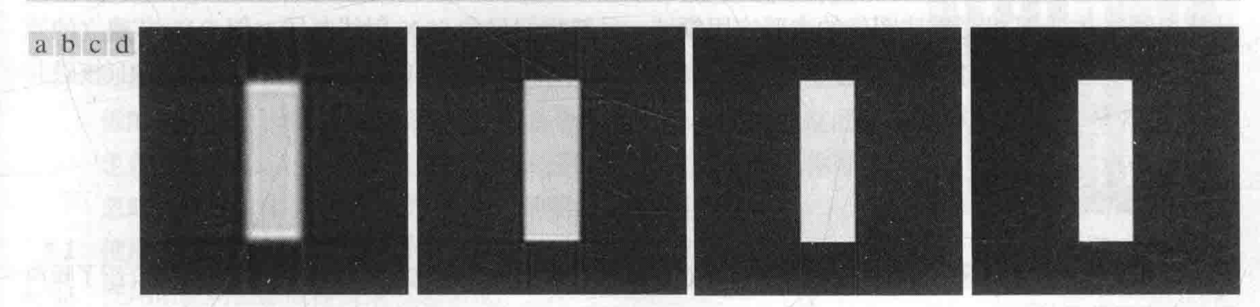


图 5.48　使用滤波扇形反投影的矩形图像的重建：(a) α 和 β 的增量为1°的结果；(b)增量为0.5°的结果；(c)增量为0.25°的结果；(d)增量为0.125°的结果。请比较图(d)和图5.43(b)

除了如正弦干扰那样的混淆更为可见外，类似的结果使用头部幻影也可以得到。在图5.49(c)中，我们看到，即使使用 $\Delta\alpha = \Delta\beta = 0.25°$，严重的失真也依然存在，特别是在椭圆的外围。使用矩形时，使用0.125°的增量最终产生了可与图5.44(b)中头部幻影反投影图像相比的结果。这些结果说明了在现代CT系统的扇形射线几何原理中，必须使用数以千计的检测器来减少混淆效应的原因。

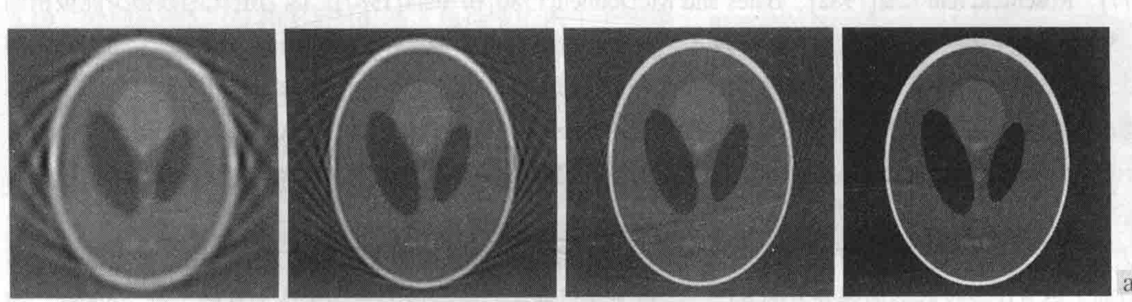


图 5.49　使用滤波扇形反投影的头部幻影图像的重建：(a) α 和 β 的增量为1°的结果；(b)增量为0.5°的结果；(c)增量为0.25°的结果；(d)增量为0.125°的结果。请比较图(d)和图5.44(b)

407~409

小结

本章中的复原结果基于这样一个假设，即图像退化可建模为一个线性的、位置不变的过程，以及与图像值不相关的加性噪声。即使这些假设并不完全正确，用前几章提出的方法得到有用的结果通常也是可能的。

本章推导的某些复原技术基于各种最佳准则。“最佳”一词的使用涉及严格的数学概念，而不是指人类视觉系统的最佳响应。事实上，对于视觉感知知识的缺乏妨碍了图像复原问题的一般表达，而图像复原问题要考虑到观察者的偏好与能力。在这些限制条件下，本章所介绍概念的优势是基本方法的推导，它有合理的预测能力和坚实的基础知识的支持。

就像在第3章和第4章中那样，一定的复原任务，譬如降低随机噪声，是在空间域中使用卷积模板来执行的。我们还发现频率域对于降低周期性噪声和对于某些重要的退化建模是很理想的，如在图像获取期间因为运动导致的模糊。我们还发现，对于表达复原滤波器来说，频率域也是很有用的工具，如维纳滤波器和约束最小二乘方滤波器。

如第4章中提到的那样，频率域为实验提供了直观且可靠的基础。一旦为给定的应用找到了令人满意的一种方法(滤波器)，其实现通常就是通过数字滤波器的设计在频率域近似解决来执行的。正如第4章末尾指出的那样，在计算机中或专用的硬件/固件系统中运行速度要快得多。

投影重建图像的论述是图像处理在这一领域的基础，虽然它只是介绍性的。如5.11节说明的那样，计算机断层(CT)是投影重建图像的主要应用领域。虽然我们只关心X射线断层，但5.11节建立的原理也适用于其他CT成像模式，譬如SPECT(单光子发射断层)、PET(正电子发射断层)、MRI(核磁共振成像)和某些超声成像模式。

参考文献

关于5.1节中线性退化模型的其他读物，有Castleman[1996]和Pratt[1991]。Peebles[1993]概要介绍了噪声概率密度函数及其特性(见5.2节)，而Papoulis[1991]则深入且详细地介绍了这些概念。5.3节的参考文献有Umbaugh[2005]、Boie and Cox[1992]、Hwang and Haddad[1995]和Wilburn[1998]。自适应中值滤波的资料见Eng and Ma[2001, 2006]。自适应滤波器设计的一般领域，是5.3节讨论的自适应滤波器的优秀背景资料，Haykin[1996]详细介绍了这一主题。5.4节中的滤波器是第4章内容的延伸。5.5节内容的其他读物见Rosenfeld and Kak[1982]和Pratt[1991]。

估计退化函数这一主题(见5.6节)是当前的一个重要领域。估计退化函数的一些早期技术见Andrews and Hunt[1977]、Rosenfeld and Kak[1982]、Bates and McDonnell[1986]和Stark[1987]。因为退化函数很少能被精确地知道，所以过去几年中人们提出了大量的技术，其中强调了特殊复原。例如，Geman and Reynolds[1992]和Hurn and Jennison[1996]提出了保护灰度急剧过渡来强调清晰度的问题，Boyd and Meloche[1998]则专注于复原退化图像中的细小物体。处理模糊图像的技术的例子见Yitzhaky et al.[1998]、Harikumar and Bresler[1999]、Mesarovic[2000]和Giannakis and Heath[2000]。序列图像的复原也是考虑的重点，Kokaram[1998]为这一领域提供了很好的基础。

过去几年来，对于5.7节到5.10节讨论的滤波方法，图像处理方面的书籍和论文中已用各种方法进行了解释。两种主要方法是这些滤波器开发的基础。一种方法以使用矩阵理论的通用表达为基础，见Andrews and Hunt[1977]。这种方法精致且通用，但对于新涉猎该领域的人来说比较困难，因为它缺乏直观性。直接基于频率域滤波的方法(本章采用的方法)对于初次接触复原的人来说，通常比较容易，但缺少矩阵法统一的数学严谨

409 ~ 410

性。两种方法都可以达到相同的结果，但根据我们的授课经验，初涉这一领域的学生更喜欢后一种方法。5.7 节到 5.10 节所述滤波概念的补充读物是 Castleman[1996]、Umbaugh[2005]和 Petrou and Bosdogianni [1999]，其中后者很好地关联了二维频域滤波器和相应的数字滤波器。关于二维数字滤波器的设计，请参阅 Lu and Antoniou[1992]。

计算机断层的基本参考资料是 Rosenfeld and Kak[1982]、Kak and Slaney[2001]和 Prince and Links[2006]。关于 Shepp-Logan 幻影的参考资料见 Shepp and Logan[1974]，关于 Ram-Lak 滤波器起源的详细内容见 Ramachandran and Lakshminarayanan[1971]。O'Connor and Fessler[2006]是有关计算机断层在信号与图像处理领域当前研究的代表性论文。

本章中讨论的多数内容的软件实现，请参阅 Gonzalez, Woods and Eddins[2004]。

标有星号的习题的详细解答可在本书的网站上找到。该网站还包含了基于本章内容所建议的项目。

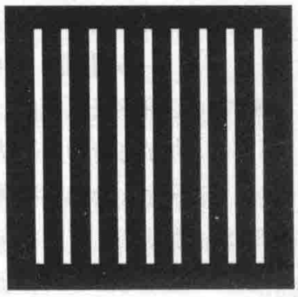
习题

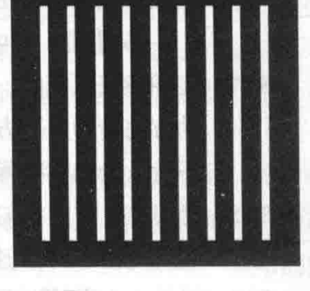
★**5.1** 测试模式中白色条带的尺寸为 7 像素宽、210 像素高。两个白色条带之间的间距为 17 像素。应用下面的处理后，该图像看起来像什么？

(a) 3×3 算术均值滤波器。

(b) 5×5 算术均值滤波器。

(c) 9×9 算术均值滤波器。





注意：与对该图像滤波相关的这一习题及后面几个习题看起来有点单调，但它们值得尝试，因为它们可帮助我们理解这些滤波器如何工作。理解了一个特殊滤波器对图像的影响后，你的回答可以是关于结果图像的简短描述。例如，“结果图像将由 3 个像素宽、206 个像素高的白色条带组成”。一定要描述白色条带的任何变形，例如圆角。你可以忽略模板仅包含了部分图像像素引起的图像边界效应。

5.2 使用几何均值滤波器重做习题 5.1。

★**5.3** 使用谐波均值滤波器重做习题 5.1。

5.4 使用逆谐波均值滤波器($Q = 1.5$)重做习题 5.1。

★**5.5** 使用逆谐波均值滤波($Q = -1.5$)重做习题 5.1。

5.6 使用中值滤波器重做习题 5.1。

★**5.7** 使用最大值滤波器重做习题 5.1。

5.8 使用最小值滤波器重做习题 5.1。

★**5.9** 使用中点滤波器重做习题 5.1。

5.10 分别从图5.7(c)和图5.7(d)右上角提取的两个子图像如右所示。左侧的子图像是使用大小为 3×3 的算术均值滤波器处理的结果，右侧的子图像是使用相同大小的几何均值滤波器处理的结果。

★(a) 解释为什么使用几何均值滤波得到的子图像模糊较少。（提示：你的分析可通过检查灰度的一维阶跃过渡来开始。）

(b) 解释右侧图像中黑色元件更粗的原因。

5.11 根据式(5.3-6)给出的逆谐波滤波器，回答如下问题。

(a) 解释为什么当 Q 为正值时，该滤波器可有效去除“胡椒”噪声。

410~412

(b) 解释为什么当 Q 为负值时，该滤波器可有效去除“盐粒”噪声。

(c) 解释为什么选择的 Q 值不当时，该滤波器给出较差结果(譬如图 5.9 所示的结果)。

(d) 讨论当 $Q=-1.5$ 时该滤波器的特性。

(e) 讨论(对于 Q 为正和负的情况)恒定灰度级区域中该滤波器的特性。

★5.12 给出与表 4.6 中带阻滤波器对应的高斯和布特沃斯带通滤波器的公式。

5.13 以式(4.10-5)的形式给出高斯、布特沃斯和理想陷波带阻滤波器的公式。

★5.14 证明二维连续余弦函数

$$f(x,y)=A\cos(u_0x+v_0y)$$

的傅里叶变换是共轭冲激对

$$f(u,v)=-\frac{A}{2}\left[\delta\left(u-\frac{u_0}{2\pi},v-\frac{u_0}{2\pi}\right)+\delta\left(u+\frac{u_0}{2\pi},v+\frac{u_0}{2\pi}\right)\right]$$

[提示：使用式(4.5-7)中的连续傅里叶变换，并以指数形式表示该余弦函数。]

5.15 使用式(5.4-11)推导式(5.4-13)。

★5.16 考虑一个线性、位置不变的图像退化系统，其冲激响应为

$$h(x-\alpha,y-\beta)=\mathrm{e}^{-[(x-\alpha)^2+(y-\beta)^2]}$$

假设系统的输入是由宽度无限小的一条线组成的图像，这条线位于 $x=a,y=b$ 处并由 $f(x,y)=\delta(x-a,y-b)$ 建模，其中 δ 是一个冲激函数。假设无噪声，输出图像 $g(x,y)$ 是什么？

5.17 在图像获取期间，在时间 T_1 内，一幅图像在垂直方向上做均匀线性运动。然后，在时间 T_2 内，运动方向转为水平方向。假设图像改变方向所用的时间可以忽略不计，且快门开关时间同样可以忽略不计，给出模糊函数 $H(u,v)$ 的表达式。

★5.18 考虑在 x 方向均匀加速导致的图像模糊问题。如果图像在时刻 $t=0$ 时静止，在时间 T 内，以均匀加速度 $x_0(t)=at^2$ 加速，请找出模糊函数 $H(u,v)$。可以假设快门开关时间忽略不计。

5.19 设计一个在接近着陆时从行星发送图像的空间探测器。在着陆的最后阶段，一个控制推进器坏了，导致探测器沿其垂直轴急速旋转。在着陆期间的最后两秒发送的图像由于这种圆周运动而被模糊。摄像机位于探测器的底部，沿探测器的垂直轴并指向下方。所幸的是，探测器的转动也是沿着垂直轴的，因此图像是由均匀的旋转运动模糊的。在每幅图像获取期间，探测器旋转 $\pi/12$ 弧度。假设在图像获取期间垂直运动可以忽略。请以公式形式给出复原图像的解决方案。

★5.20 如右所示的图像是心脏立体再现的一个模糊的二维投影。已知图像右侧底部的每个十字线宽 4 像素、长20像素，且模糊之前的灰度值为255。试给出一步一步的过程，指出你怎样应用刚才提供的信息得到模糊函数 $H(u,v)$。



(原图像由 GE 医学系统提供)

5.21 某 X 射线成像几何产生了一种模糊退化，这种模糊退化可用空间圆对称函数

$$h(x,y)=\frac{x^2+y^2-\sigma^2}{\sigma^4}\mathrm{e}^{-\frac{x^2+y^2}{2\pi^2}}$$

与该遥感图像的卷积来建模。假设是连续变量，证明频率域中的退化由下式给出：

$$H(u,v)=-\sqrt{2\pi}\sigma(u^2+v^2)\mathrm{e}^{-2\pi^2\sigma^2(u^2+v^2)}$$

(提示：从 4.9.4 节、表 4.3 中的性质 13 和习题 4.26 入手。)

412 ~ 414

★5.22 使用习题 5.21 中的传递函数，给出几何均值滤波器的表达式，假设噪声功率谱和未退化信号之比为常数。然后，给出维纳滤波器的表达式。

5.23 使用习题 5.21 中的传递函数，给出约束最小二乘方滤波器的最终表达式。

5.24 假设图 5.1 中的模型是线性且位置不变的，并且噪声与图像是不相关的。证明输出功率谱为

$$|G(u,v)|^2 = |H(u,v)|^2|F(u,v)|^2 + |N(u,v)|^2$$

参考式(5.5-17)和式(4.6-18)。

5.25 Cannon[1974]建议过一个满足条件

$$|\hat{F}(u,v)|^2 = |R(u,v)|^2|G(u,v)|^2$$

的复原滤波器 $R(u, v)$，并且以强制复原图像的功率谱 $|\hat{F}(u,v)|^2$ 等于原图像的功率谱 $|F(u, v)|^2$ 为前提。假设该图像和噪声是不相关的。

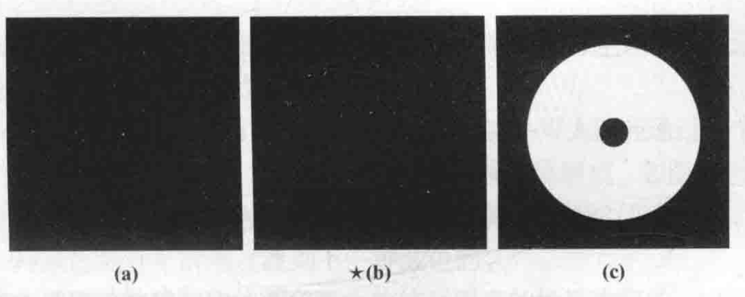
★(a) 根据 $|F(u,v)|^2$、$|H(u,v)|^2$ 和 $|N(u,v)|^2$ 给出 $R(u, v)$。［提示：参阅图 5.1、式(5.5-17)和习题 5.24］。

(b) 使用(a)中得到的结果来说明式(5.8-2)给出的结果。

5.26 一位天文学家使用一台高倍望远镜观察到的图像有点模糊。制造商告诉天文学家，设备是按说明书操作的。望远镜镜头把图像聚焦到高分辨率 CCD 成像阵列上，然后望远镜使用电子学方法将图像转换为数字图像。试图通过传导控制实验室的实验使用镜头和成像传感器来改进这种情况是不可能的，因为望远镜部件的尺寸和重量太大。天文学家听说你是一位成功的图像处理专家，打电话要求你帮助她提出将图像变得清晰一些的数字图像处理方案。假设你能得到的图像仅仅是星体图像，请问你如何解决该问题?

★5.27 一位考古学教授正在从事罗马帝国时期流通货币方面的研究工作。他最近了解到对其研究起决定性作用的 4 个罗马硬币陈列在伦敦的大英博物馆中。遗憾的是，在到达博物馆后，他被告知这些硬币最近被盗了。进一步的研究表明，依靠博物馆保存的这些硬币的照片也可行。遗憾的是，硬币的照片模糊为一些点，以至于无法识别数据和小标记。模糊的原因在于拍摄照片时摄像机散焦。作为一名图像处理专家和教授的朋友，要求你帮助决定计算机处理是否能用于复原图像，并为教授指出在哪里能读出这个标记。而且你被告知，与同一时期其他有代表性的硬币一样，用于拍摄这些照片的摄像机仍然可用。提出解决这一问题的一步一步的过程。

5.28 画出下列方形图像的雷登变换。在画出的图形中定量地标出所有的重要特性。图(a)由中间的一个点组成，图(b)沿对角线有两个点。根据灰度剖面来描述你对图(c)的解决方案。假设使用平行射线几何理论。



(a) ★(b) (c)

5.29 证明高斯函数 $f(x,y) = A\exp\left(\dfrac{-x^2-y^2}{2\sigma^2}\right)$ 的雷登变换［见式(5.11-3)］是 $g(\rho,\theta) = A\sqrt{2\pi}.\sigma\exp(-\rho^2)$。

［提示：参考例 5.17，这里我们使用对称性来简化积分。］

414 ~ 415

5.30★ (a) 证明单位冲激函数 $\delta(x, y)$ 的雷登变换［见式(5.11-3)］是 $\rho\theta$ 平面中通过原点的一条垂直线。

(b) 证明冲激 $\delta(x-x_0, y-y_0)$ 的雷登变换是在 $\rho\theta$ 平面中的一条正弦曲线。

5.31 证明雷登变换［见式(5.11-3)］的下列性质的正确性：

★(a) 线性：雷登变换是一个线性算子(见 2.6.2 节关于线性算子的定义)。

(b) 平移性质：$f(x-x_0, y-y_0)$ 的雷登变换是 $g(\rho - x_0\cos\theta - y_0\sin\theta, \theta)$ 。

★(c) 卷积性质：证明两个函数的卷积的雷登变换等于两个函数的雷登变换的卷积。

5.32 给出由式(5.11-14)推导出式(5.11-15)的步骤。你需要用到性质 $G(\omega, \theta+180°) = G(-\omega, \theta)$ 。

★**5.33** 证明式(5.11-25)的正确性。

5.34 证明式(5.11-27)的正确性。

第6章　彩色图像处理

> It is only after years of preparation that the young artist should touch color—not color used descriptively, that is, but as a means of personal expression.
>
> *Henri Matisse*

> For a long time I limited myself to one color—as a form of discipline.
>
> *Pablo Picasso*

引言

在图像处理中，彩色的运用主要受两个因素的推动。第一，彩色是一个强有力的描绘子，它常常可简化从场景中提取和识别目标；第二，人可以辨别几千种彩色色调和亮度，但相比之下只能辨别几十种灰度色调。第二个因素在人工图像分析中特别重要。

彩色图像处理可分为两个主要领域：全彩色处理和伪彩色处理。在第一类中，通常要求图像用全彩色传感器获取，如彩色电视摄像机或彩色扫描仪。在第二类中，问题是对一种特定的单色灰度或灰度范围赋予一种颜色。到目前为止，多数数字彩色图像处理是在伪彩色层面完成的。然而，在过去的十年里，彩色传感器和用于处理彩色图像的硬件的价格上变得更容易接受，导致全彩色图像处理技术的应用日益广泛，包括出版、可视化和互联网应用。

在后续的讨论中，我们会看到前面几章介绍的一些灰度方法可直接用于彩色图像。其他方法则要求将灰度方法重新表达为与本章推导的彩色空间的性质保持一致。本书讨论的这些技术远未穷尽，仅是用于彩色图像处理的一些方法。

6.1　彩色基础

虽然人的大脑感知和理解颜色所遵循的过程还是未完全了解的一种生理与心理现象，但颜色的物理性质可由实验和理论结果支持的基本形式来表示。

1666年，艾萨克·牛顿(Isaac Newton)发现了一种现象，当一束太阳光通过一个玻璃棱镜时，出现的光束不是白色的，而是由一端为紫色、另一端为红色的连续色谱组成。如图6.1所示，色谱可分为6个较宽的区域：紫色、蓝色、绿色、黄色、橙色和红色。观察全彩色时(见图6.2)，色谱末尾的颜色不是突变的，而是从一种颜色混合平滑地过渡到下一种颜色的。

基本上，人类和某些其他动物感知物体颜色是由物体反射光的性质决定的。如图6.2所示，可见光由电磁波谱中相对较窄的频段组成。物体反射的光若在所有可见光波长范围内是平衡的，那么对观察者来说显示为白色。然而，一个物体反射有限的可见光谱时，则呈现某种颜色。例如，绿色物体反射具有500~570 nm范围内主要波长的光，吸收其他波长的多数能量。

416~417

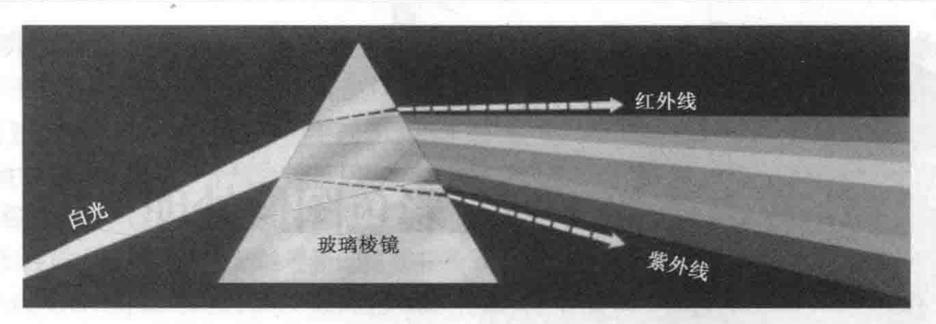


图 6.1 白光通过棱镜看到的色谱(原图像由 General Electric 公司的 Lamp Business 分公司提供)

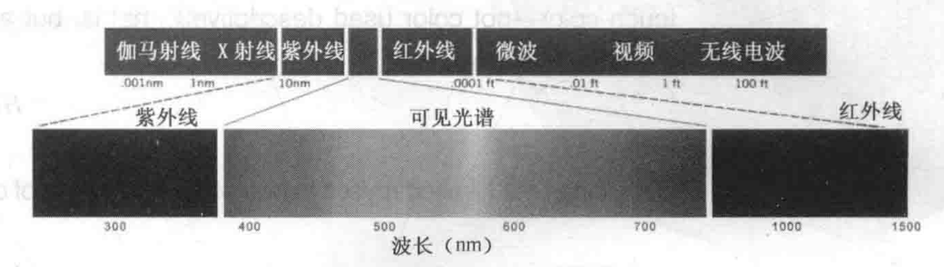


图 6.2 可见范围电磁波谱的波长组成(原图像由 General Electric 公司的 Lamp Business 分公司提供)

光的特性是色彩学的核心。如果光是无色的(缺乏颜色)，那么其属性仅是亮度或数值。无色光就是观察者在黑白电视机上看到的光，且它是迄今为止我们对图像处理的讨论的隐含分量。如第 2 章中定义并多次使用的那样，术语灰度级仅提供了一个亮度的标量度量，它的范围从黑色到灰色最终到白色。

彩色光大约覆盖电磁波谱中 400~700 nm 的范围。用来描述彩色光源质量的 3 个基本量是辐射、光强和亮度。辐射是从光源流出的能量的总量，通常用瓦特(W)来度量。光强用流明来度量，它给出观察者从光源感知的能量总和的度量。例如，由远红外波谱范围的光源发出的光，可能具有很大的能量(辐射)，但观察者却很难感知到它；它的光强几乎是零。最后，亮度是一个主观描绘子，它实际上是不可度量的。它体现了无色的强度的概念，并且是描述色彩感觉的一个关键因素。

如 2.1.1 节中解释的那样，人眼中的锥状细胞是负责彩色视觉的传感器。详细的实验结果已经确定，人眼中的 600~700 万个锥状细胞分为 3 个主要的感知类别，分别对应于红色、绿色和蓝色。大约 65%的锥状细胞对红光敏感，33%对绿光敏感，只有 2%对蓝光敏感(但蓝色锥状细胞对蓝光更敏感)。图 6.3 显示了人眼中的红色、绿色和蓝色锥状细胞吸收光的平均实验曲线。由于人眼的这些吸收特性，所看到的彩色是所谓的原色红(R)、绿(G)、蓝(B)的各种组合。为标准化起见，CIE(国际照明委员会)于 1931 年指定了如下特定波长值为三原色：蓝色 = 435.8 nm，绿色 = 546.1 nm，红色 =700 nm。这一标准在图 6.3 所示的详细实验曲线于 1965 年被采用之前就做了规定。因此 CIE 标准只是近似对应于实验数据。由图 6.2 和图 6.3 我们注意到，没有单一的颜色可称为红色、绿色或蓝色。另外，应记住为标准化目的而定的三个特定原色波长，并不意味着这三个固定的 RGB 分量单独作用就能产生所有谱，这一点很重要。“原色”一词的使用已被广泛误解为：三个标准原色以各种强度比混合在一起时，可以产生所有的可见彩色。如我们很快将会看到的那样，这一解释是不正确的，除非波长允许变化，而在这种情况下将不再有三个固定的标准原色。

原色相加可以产生二次色，如深红色(红色加蓝色)、青色(绿色加蓝色)和黄色(红色加绿色)。以正确的亮度把三原色或把与二次色相对应的原色混合，即可产生白光。这一结果示于图 6.4(a)，它还说明了三原色及其混合产生二次色。

417~418

光的原色与颜料或着色剂的原色之间的区别很重要。后者的原色定义为减去或吸收光的一种

原色，并反射或传输另外两种原色。因此，颜料的原色是深红色、青色和黄色，而二次色是红色、绿色和蓝色。这些颜色示于图 6.4(b)。将三种颜料原色或与二次色对应的原色适当混合，即可产生黑色。

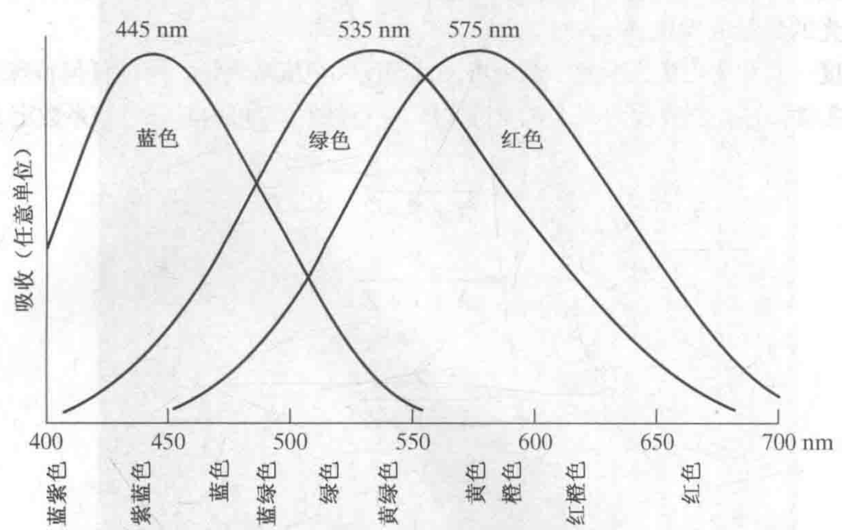


图 6.3　以波长为函数，人眼中的红色、绿色和蓝色锥状细胞对光的吸收曲线

彩色电视接收机是光的彩色相加性质的一个例子。许多彩色电视荧光屏(CRT，阴极射线管)的内部是由电敏荧光粉按三角形点阵的模式组成的。当激发时，三色组中的每一点产生三原色中的一种光。发红光的荧光点的亮度由显像管内的电子枪调制，电子枪产生的脉冲与电视摄像机看到的“红色能量”相对应。每个三色组中的绿色荧光点和蓝色荧光点以相同的方式调制。在电视接收机上观察到的效果是来自每个三色组的三原色“加”在一起，并被人眼中对颜色敏感的锥状细胞以全彩色图像的方式接收。以所有这三种颜色每秒连续变化 30 幅图像，在荧光屏上即可完成连续图像的显示。

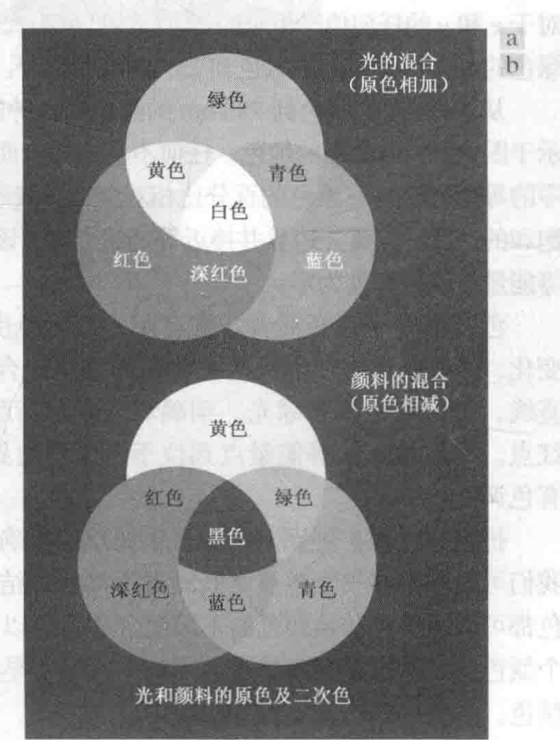


图 6.4　光和颜料的原色及二次色(原图像由 General Electric 公司的 Lamp Business 分公司提供)

CRT 显示器正在被平板数字技术取代，譬如液晶显示器(LCD)和等离子设备。虽然它们的基本原理与 CRT 不同，但在某种意义上，这些技术及类似的技术使用了一个相同的原理，即它们都要求三个子像素(红色、绿色和蓝色)去产生单个彩色像素。LCD 使用极化光的特性来阻止光或使光通过 LCD 屏幕，在有源矩阵显示技术的情况下，使用薄膜晶体管(TFT)提供适当的信号对屏幕上的每个像素寻址。光滤波器用于在每个像素三色组位置上产生光的三原色。在等离子装置中，像素是用于产生三原色之一的涂有荧光粉的微小气体单元。各个小单元以类似于 LCD 的方式寻址。这种三像素组的同等寻址能力是数字显示的基础。

418～420

通常用以区别不同颜色特性的是亮度、色调和饱和度。正如本节前面指出的那样，亮度具体表

达了无色的强度概念。色调是光波混合中与主波长有关的属性。色调表示观察者感知的主要颜色。这样，当我们说一个物体为红色、橙色或黄色时，指的是其色调。饱和度指的是相对纯净度，或一种颜色混合白光的数量。纯谱色是全饱和的。如深红色(红加白)和淡紫色(紫加白)这样的彩色是欠饱和的，饱和度与所加白光的数量成反比。

色调与饱和度一起称为色度，因此，颜色可用其亮度和色度来表征。形成任何特殊彩色的红、绿、蓝的数量称为三色值，并分别表示为X, Y和Z。这样，一种颜色就可由其三色值系数定义为

$$x = \frac{X}{X+Y+Z} \tag{6.1-1}$$

$$y = \frac{Y}{X+Y+Z} \tag{6.1-2}$$

和

$$z = \frac{Z}{X+Y+Z} \tag{6.1-3}$$

从以上公式可得①

$$x + y + z = 1 \tag{6.1-4}$$

对可见光谱内光的任何波长，用于产生对应于该波长的颜色所需的三色值，可直接从大量实验结果编制的曲线或表中得到(Poynton[1996])，也可参阅早期的文献 Walsh[1958]和 Kiver[1965]。

确定颜色的另一种方法是使用CIE色度图(见图6.5)，该图以x(红)和y(绿)的函数表示颜色的组成。对于x和y的任何值，相应的z值(蓝色)可由式(6.1-4)得到，注意$z = 1 - (x + y)$。例如，图6.5中标记为绿色的点大约有62%的绿色和25%的红色成分。从式(6.1-4)可得到蓝色的成分约为13%。

从380 nm的紫色到780 nm的红色的各种谱色的位置，标在舌形色度图周围的边界上。这些都是示于图6.2的谱图中的纯色。任何不在边界上而在色度图内部的点，都表示谱色的混合色。示于图6.5中的等能量点与三原色的百分比相对应，它表示白光的CIE标准。位于色度图边界上的任何点都是全饱和的。当一点离开边界并接近等能量点时，该颜色中就加入了更多的白光，它就变成欠饱和的颜色。等能量点的饱和度为零。

色度图对于色彩混合非常有用，因为色度图中连接任意两点的直线段定义了所有不同颜色的变化，这些颜色可以由这两种颜色的加性组合得到。例如，考虑图6.5中从红点到绿点画出的一条直线，如果红光多于绿光，则确切地表示新颜色的点将处在该线段上，但与绿点相比它更接近于红点。类似地，从等能量点到位于色度图边界上的任意一点画一线段，将定义那个特定谱色的所有色调。

把这一过程扩展到三种颜色很简单。要确定从色度图中任何给定的三种颜色得到的颜色范围，我们可简单地在三个色点之间画出连接线，结果是一个三角形，三角形边界上或三角形内的任何颜色都可以由三种初始颜色的不同组合产生。以任意固定颜色为顶点的三角形无法包围图6.5中的整个颜色区域。这种观察结果从图解上支持了早期的结论，即用3个单一且固定的原色无法得到所有颜色。

图6.6中的三角形显示了由RGB监视器产生的典型颜色范围(称为彩色全域)。三角形内的不规则区域是今天高质量彩色打印设备的代表性彩色域。彩色打印设备的彩色域的边界是不规则的，因为彩色打印是相加色彩和相减色彩混合的组合，与在监视器上显示颜色(基于三种高度可控的原色光的相加)相比，这是一个更加难以控制的过程，监视器是以三个高度可控的原色光的相加为基础的。

420 ~ 422

① 这段文字中使用的x, y, z遵循国际惯例。这些不应与书中其他小节用于表示空间坐标的(x, y)相混淆。

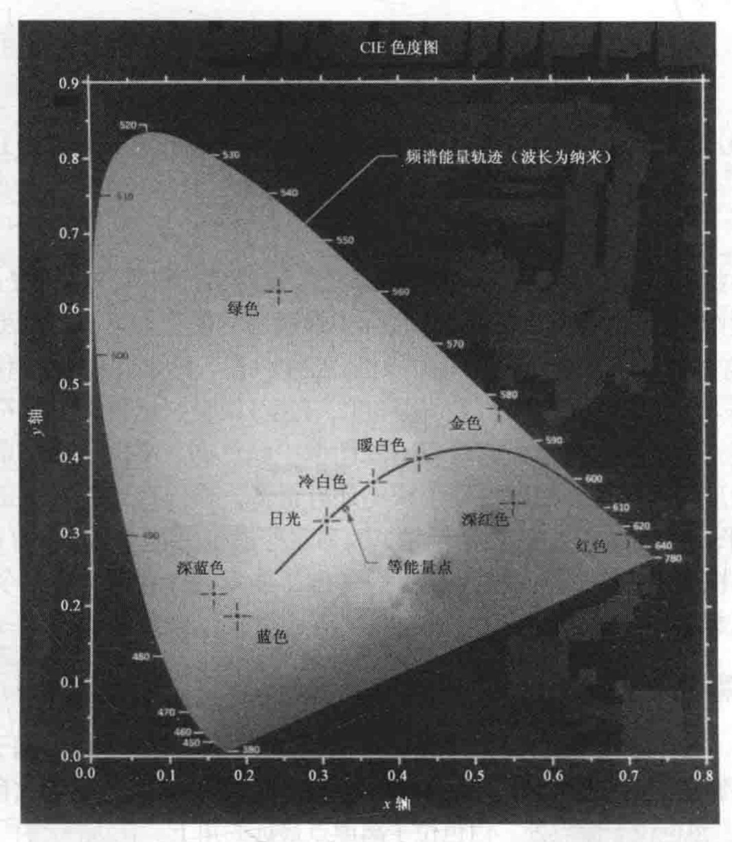


图 6.5 色度图(原图像由 General Electric 公司的 Lamp Business 分公司提供)

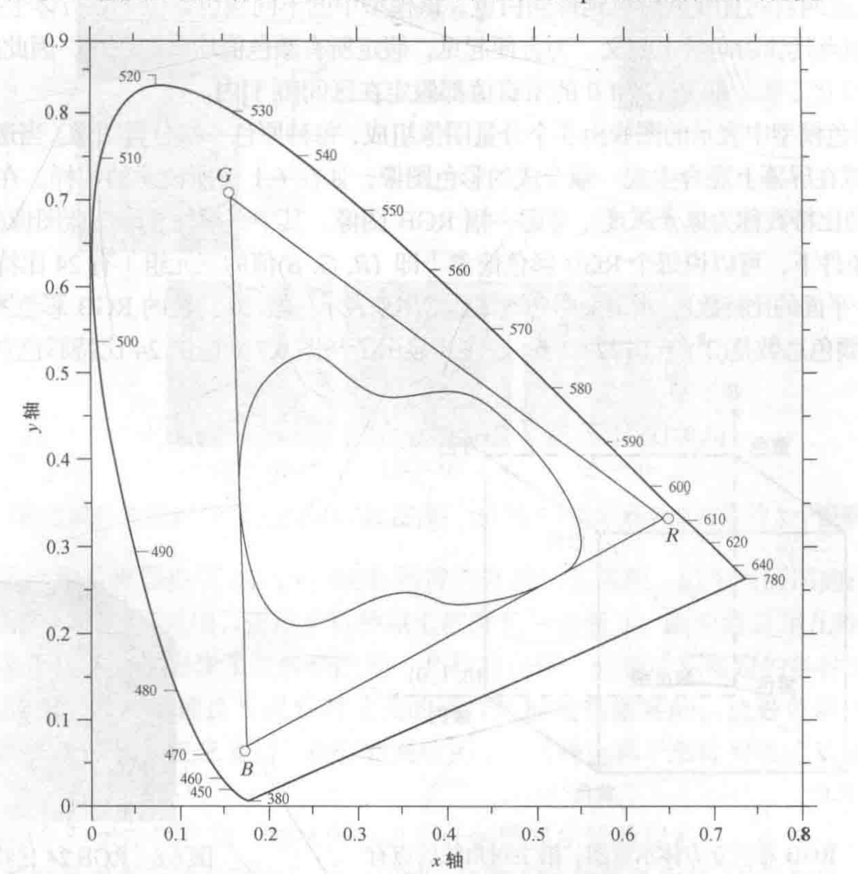


图 6.6 彩色监视器的典型彩色域(三角形区域)和彩色打印设备的典型彩色域(不规则区域)

422 ~ 423

6.2 彩色模型

彩色模型(也称为彩色空间或彩色系统)的目的是，在某些标准下用通常可以接受的方式方便地对彩色加以说明。本质上，彩色模型是坐标系统和子空间的说明，其中，位于系统中的每种颜色都由单个点来表示。

现在所用的大多数彩色模型不是面向硬件(如彩色监视器和打印机)的，就是面向应用的(例如针对动画的彩色图形创作)。在数字图像处理中，实际中最通用的模型如下：面向硬件的 RGB(红、绿、蓝)模型，用于彩色监视器和一大类彩色视频摄像机；CMY(青、粉红、黄)模型和 CMYK(青、粉红、黄、黑)模型，用于彩色打印机；HSI(色调、饱和度、亮度)模型，这种模型更符合人描述和解释颜色的方式。HSI 模型还有另一个优点，它可以解除图像中颜色和灰度信息的联系，使其更适合本书中给出的许多灰度处理技术。现在使用的彩色模型还有很多，原因在于色彩学是一个包括许多应用的宽泛领域。这里试图详细研究其中的几个模型，因为这些模型更有意义且更有益。然而，根据现有的任务，本章讨论的模型是图像处理的主要模型。掌握了本章的内容后，理解今天所用的其他彩色模型时就不会遇到困难。

6.2.1 RGB 彩色模型

在 RGB 模型中，每种颜色出现在红、绿、蓝的原色光谱成分中。该模型基于笛卡儿坐标系。所考虑的彩色子空间是图6.7所示的立方体，图中 RGB 原色值位于 3 个角上；二次色青色、深红色和黄色位于另外 3 个角上，黑色位于原点处，白色位于离原点最远的角上。在该模型中，灰度(RGB 值相等的点)沿着连接这两点的直线从黑色延伸到白色。该模型中的不同颜色是位于立方体上的或立方体内部的点，且由自原点延伸的向量来定义。为方便起见，假定所有颜色值均已归一化，因此图 6.7 所示的立方体是一个单位立方体，即 R, G 和 B 的所有值都假定在区间[0, 1]内。

在 RGB 彩色模型中表示的图像由 3 个分量图像组成，每种原色一幅分量图像。当送入 RGB 监视器时，这 3 幅图像在屏幕上混合生成一幅合成的彩色图像，如在 6.1 节所说明的那样。在 RGB 空间中，表示每个像素的比特数称为像素深度。考虑一幅 RGB 图像，其中每幅红、绿、蓝图像都是一幅 8 比特图像，在这种条件下，可以说每个 RGB 彩色像素［即 (R, G, B)值的三元组］有 24 比特的深度（3 个图像平面乘以每个平面的比特数）。术语全彩色图像通常用来表示一幅 24 比特的 RGB 彩色图像。在 24 比特 RGB 图像中，颜色总数是$(2^8)^3 = 16\,777\,216$。图 6.8 显示了与图 6.7 对应的 24 比特彩色立方体。

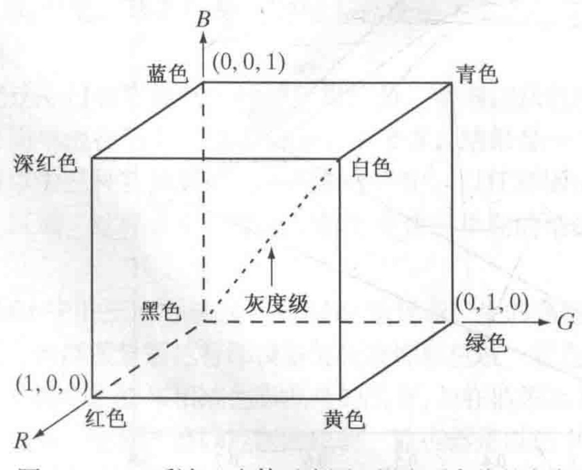


图 6.7 RGB 彩色立方体示意图。沿主对角线的点有从原点的黑色至点(1, 1, 1)的白色的灰度值

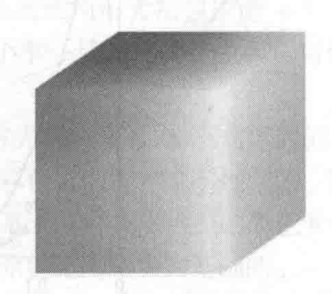


图 6.8 RGB 24 比特彩色立方体

423
~
425

例 6.1　生成 RGB 彩色立方体的隐藏面和剖面。

图6.8所示的立方体是实心的，它由前一段中提到的$(2^8)^3$ = 16 777 216 种颜色组成。观察这些颜色的一种简便方法是，生成一个彩色平面(立方体的表面或剖面)。这是通过固定三种颜色中的一种并让其他两种颜色变化来完成的。例如，在图 6.8 中，通过立方体中心并与 GB 平面平行的剖面是平面$(127, G, B)$，其中 $G, B = 0, 1, 2, \cdots, 255$。这里，我们使用了实际像素值，而不是为数学上的方便而在[0, 1]区间内的归一化值，因为前者是在计算机中实际用来产生颜色的值。图 6.9(a)说明，简单地把三幅独立的分量图像送入彩色监视器，可观察一幅横截面图像。在分量图像中，0 代表黑色，255 代表白色(注意这些是灰度级图像)。最后，图6.9(b)显示了以同样方式生成的图 6.8 中立方体的 3 个隐藏面。

值得注意的是，获取一幅彩色图像基本上是图 6.9 所示的相反过程。使用分别对红、绿、蓝敏感的 3 个滤色片可获取一幅彩色图像。当我们采用配备有这些滤色片之一的单色摄像机观察彩色场景时，结果是一幅单色图像，其亮度与滤色片的响应成正比。用每个滤色片重复这一过程，可产生三幅单色图像，这些图像就是彩色场景的 RGB 分量图像(实际上，RGB 彩色图像传感器通常将这一过程集成在一个装置中)。很明显，以图 6.9(a)所示的形式显示这 3 幅 RGB 分量图像，就会产生原始彩色场景的 RGB 彩色复现。

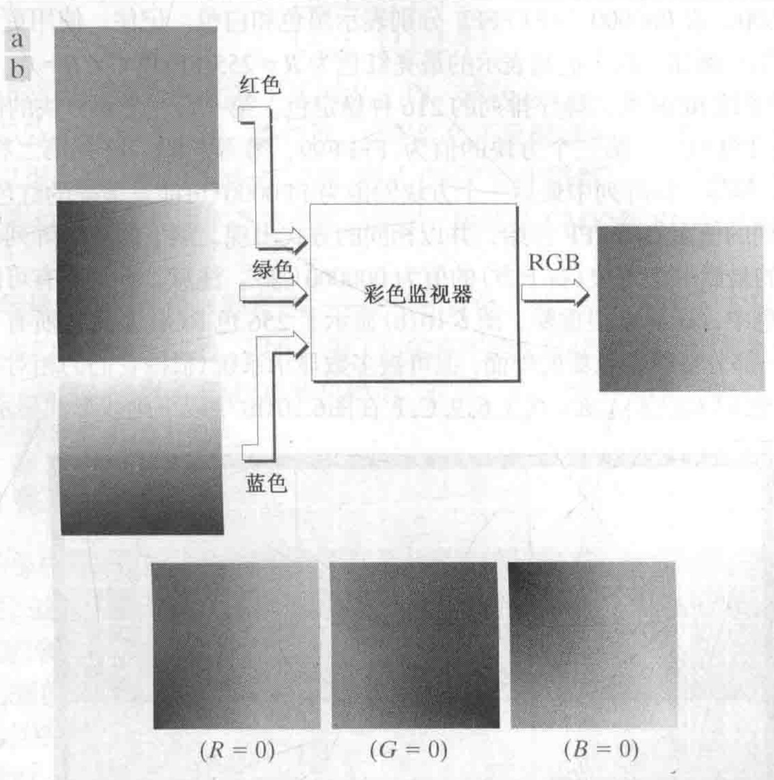


图 6.9　(a)生成彩色剖面$(127, G, B)$的 RGB 图像；(b)图 6.8 所示彩色立方体的 3 个隐藏面

虽然高端显示卡和监视器提供 24 比特 RGB 图像的合理彩色再现，但今天所用的许多系统仍限于 256 种颜色。另外，有许多应用，使用几百种以上的颜色没有意义，而有时只用几种颜色。6.3 节讨论的伪彩色图像处理技术就提供了这种情况的一个较好例子。当前应用给定的各种系统中，有一种彩色子系统值得考虑，其合理且与观察者无关的硬件性能是很重要的。这种色彩子集称为稳定 RGB 色集合，或称为全系统稳定色集合。在互联网应用中，这种色彩子集称为稳定 Web 色或稳定浏览器色。

假定 256 种颜色是最小颜色数，这些颜色可用任何能显示所希望彩色的系统真实再现，那么使用可接受的标准符号来表示这些颜色是很有用的。已知 256 种颜色中的 40 种颜色可被各种操作

425~426

系统进行不同的处理，剩下的 216 种颜色是多数系统通用的。这 216 种颜色已成为事实上的稳定色，特别是在互联网中的应用。无论在何时应用，希望大多数人观察到的颜色都一样。

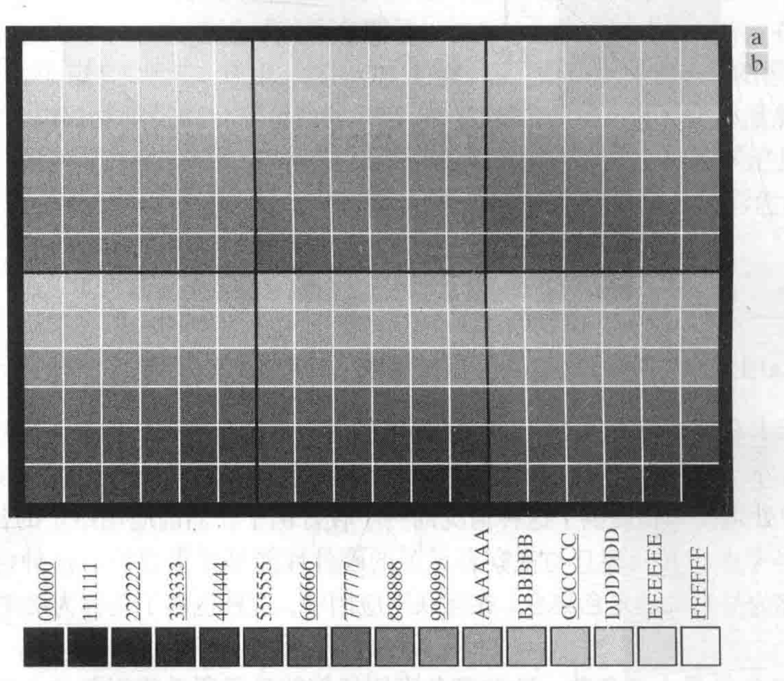
如以前一样，216 种稳定色中的每种都可以由 3 个 RGB 值形成，但每个值只能是 0, 51, 102, 153, 204 或 255。这样，这些 RGB 三元组数值就可提供$(6)^3 = 216$种可能的值(注意所有值都可被 3 整除)。通常这些值可用十六进制数来表示，如表 6.1 所示。回顾可知，十六进制数 0, 1, 2, ⋯, 9, A, B, C, D, E, F 对应十进制数 0, 1, 2, ⋯, 9, 10, 11, 12, 13, 14, 15。还可以表示为$(0)_{16} = (0000)_2$和$(F)_{16} = (1111)_2$，例如，$(FF)_{16} = (255)_{10} = (11111111)_2$，并且我们可以看到，两个十六进制数构成一个 8 比特字节。

表 6.1 稳定色中每个 RGB 分量的有效值

数 制	等效颜色					
十六进制	00	33	66	99	CC	FF
十进制	0	51	102	153	204	255

因为要取 3 个数来形成 RGB 彩色，每种稳定色由表 6.1 中的 3 个两位十六进制数形成。例如，最纯净的红色是 FF0000。值 000000 和 FFFFFF 分别表示黑色和白色。记住，使用更为熟悉的十进制表示可得到相同的结果。例如，以十进制表示的最亮红色为 $R = 255$(FF)和 $G = B = 0$。

图6.10(a)显示了按 RGB 的值降序排列的 216 种稳定色。第一行左上角方块的值为 FFFFFF(白)，其右侧方块的值为 FFFFCC，第三个方块的值为 FFFF99，等等。同一阵列第二行的值是 FFCCFF, FFCCCC, FFCC99，等等。该阵列中最后一个方块的值为 FF0000(可能是最亮的红色)。刚考察过的阵列右侧的第二个阵列的值由 CCFFFF 开始，并以相同的方式出现，剩下的 4 个阵列也以同样的方式出现。最后一个阵列的最后一个方块(右下方)的值为 000000(黑)。注意，不是所有可能的 8 比特灰色都包含在 216 种稳定色中，这一点很重要。图 6.10(b)显示了 256 色 RGB 系统中所有可能灰色的十六进制代码。这些值的一部分在稳定色集的外面，但可被多数显示系统(根据它们的相对亮度)适当地表示。来自稳定色组的灰度$(KKKKKK)_{16}$, $K = 0, 3, 6, 9$, C, F 在图6.10(b)中以下画线形式显示。



426 ~ 428

图 6.10 (a)216 种稳定 RGB 色；(b)256 色 RGB 系统中的所有灰度(部分稳定色组的灰度以下画线的形式显示)

图6.11显示了RGB稳定色立方体。与图6.8所示的实心全彩色立方体不同，图6.11中的立方体仅在表面存在有效的颜色。如图6.10(a)所示，每个平面都有36种颜色，所以稳定色立方体的整个表面被216种不同的颜色所覆盖，而这正是我们所期望的。

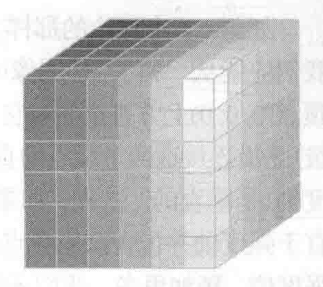


图6.11　RGB稳定色立方体

6.2.2　CMY和CMYK彩色模型

如6.1节中指出的那样，青色、深红色和黄色是光的二次色，换句话说，它们是颜料的原色。例如，当青色颜料涂覆的表面用白光照射时，该表面将不反射红光。也就是说，青色从反射的白光中减去红光，白光本身由等量的红光、绿光和蓝光组成。

大多数在纸上沉积彩色颜料的设备，如彩色打印机和复印机，要求输入CMY数据或在内部进行RGB到CMY的转换。这一转换是使用下面这个简单的操作执行的：

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{6.2-1}$$

这里再次假设所有的彩色值都归一化到了区间[0, 1]内。式(6.2-1)表明涂有青色颜料的表面所反射的光中不包含红色(即公式中$C=1-R$)。类似地，纯深红色不反射绿色，纯黄色不反射蓝色。式(6.2-1)还揭示RGB值可以很容易地通过1减去CMY值从CMY集中得到。正如早些时候指出的那样，在图像处理中，这种彩色模型主要用于产生硬拷贝输出，因此从CMY到RGB的反向操作通常没有实际意义。

根据图6.4，等量的颜料原色，即青色、深红色和黄色，可以生成黑色。实际上，为打印目的组合的这些颜色所产生的黑色是不纯的。因此，为了生成真正的黑色(即在打印中起主要作用的颜色)，加入了第4种颜色——黑色，提出了CMYK彩色模型。这样，当出版商提到“四色打印”时，指的就是CMY彩色模型的三种原色再加上黑色。

6.2.3　HSI彩色模型

如我们见到的那样，在RGB模型和CMY模型中创建颜色并从一种模型转换到另一种模型的过程很简单。如上所述，这些彩色系统对于硬件实现很理想。另外，RGB系统能很好地吻合人眼强烈感知红、绿、蓝三原色的事实。遗憾的是，RGB模型、CMY模型和其他类似的彩色模型都无法很好地适应人类实际上解释的颜色。例如，使用给定组成颜色的每种原色的百分比不能给出一辆汽车的颜色，此外，我们不认为彩色图像是由3幅原色图像混合形成的单幅图像。

观察彩色物体时，我们用其色调、饱和度和亮度来描述这个物体。回顾6.1节的讨论，色调描述的是一种纯色(纯黄色、纯橙色或纯红色)的颜色属性，而饱和度是一种纯色被白光稀释的程度的度量。亮度是一个主观描述子，实际上它是不可度量的。它体现了无色的强度概念，并且是描述彩色感觉的关键因子之一。我们清楚地知道强度(灰度级)是单色图像最有用的描述子。很明显，这个量是可度量的，并很容易解释。我们即将提出的模型称为HSI(色调、饱和度和强度)彩色模型，该模型可在彩色图像中从携带的彩色信息(色调和饱和度)中消去强度分量的影响。因此，HSI模型是开发基于彩色描述的图像处理算法的理想工具，这种彩色描述对人来说是自然且直观的，毕竟人才是这些算法的开发者和使用者。因此，我们可以概括说明如下：RGB对于图像颜色生成来说是理想的(如用彩色摄像机获取图像，或在监视器屏幕上显示图像)，但在用于颜色描述时则有许多限制。下面提供了一种这样做的有效方法。

428
~
429

如例 6.1 中讨论的那样，RGB 彩色图像可视为 3 幅单色亮度图像(表示红色、绿色和蓝色)，因此，我们应能从一幅 RGB 图像中提取出强度。如果使用图 6.7 中的彩色立方体，让白色顶点(1, 1, 1)在黑色顶点(0, 0, 0)上方直接面对它，如图 6.12(a)所示，那么这一点会更加清楚。联系图 6.7 可知，强度(灰度级)是沿连接这两个顶点的直线分布的。如图 6.12 中的排列所示，连接黑色顶点和白色顶点的直线(灰度轴)是垂直的。这样，如果我们要确定图6.12 中任何彩色点的强度分量，可以简单地找到通过一个垂直于强度轴并包含该彩色点的平面，然后找到这个平面与强度轴的交点，交点的值即为区间[0, 1]内的强度值。稍加思考，我们还注意到，一种颜色的饱和度(纯度)以强度轴的距离为函数而增大。事实上，强度轴上点的饱和度为零，事实证明沿着这条轴线的所有点都是灰度。

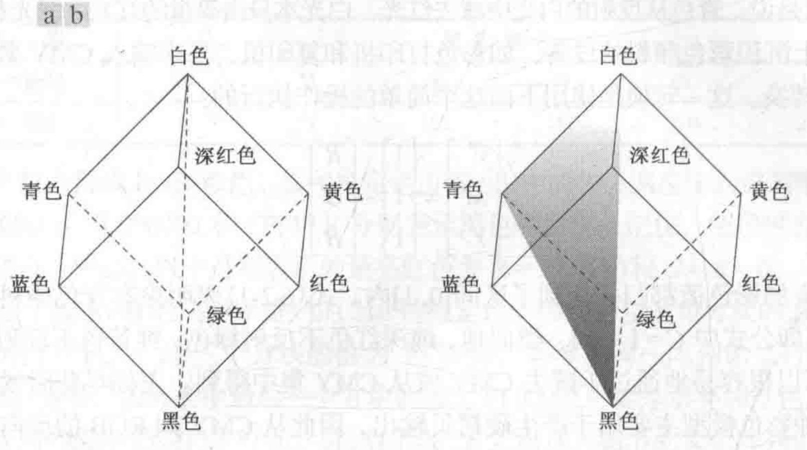


图 6.12 RGB 模型和 HSI 模型间的概念关系

要解如何从给定的 RGB 点来确定色调，我们考虑图6.12(b)，图中显示了由 3 个点(黑、白和青)定义的一个平面。黑点和白点包含在平面内的事实告诉我们，强度轴也包含在这个平面内。进一步，我们看到，包含在由强度轴和立方体边界定义的平面段内的所有点，都有相同的色调(在这种情形下为青色)。回顾 6.1 节可得到相同的结论，即所有颜色都是由位于这些颜色定义的三角形中的 3 种颜色产生的。如果这些点中的两点是黑点和白点，第三点是一个彩色点，那么三角形上的所有点都有相同的色调，因为黑分量和白分量不能改变色调(当然，这个三角形中的点的强度和饱和度是不同的)。关于垂直强度轴旋转阴影平面，可得到不同的色调。由这些概念可以得出这样的结论：形成 HSI 空间所要求的色调、饱和度和强度值可由 RGB 彩色立方体得到。也就是说，通过计算前面讨论中描述和推导的几何公式，可将任何 RGB 点转换为相应的 HSI 彩色模型中的点。

关于图6.12 中的立方体结构及其对应的 HSI 彩色空间，要记住的一个关键点是，HSI 空间由一个垂直强度轴和位于与该轴垂直的平面内的彩色点的轨迹表示。当平面沿强度轴上下移动时，由每个平面与立方体表面构成的横截面定义的边界不是三角形就是六边形。这种情况如图 6.13(a)所示，沿其灰度轴向下看，这一点就会清楚得多。在这个平面中，我们可以看到原色按 120°分隔，二次色与原色相隔 60°，这意味着二次色之间也相隔 120°。图 6.13(b)显示了相同的六边形和任意一个彩色点(显示为一个点)。该点的色调由来自某参考点的一个角度决定。通常(并非总是如此)，与红轴的 0° 角指定为 0 色调，从这里开始色调逆时针增长。饱和度(距垂直轴间的距离)是从原点到该点的向量长度。注意，原点是由该彩色平面的横截面与垂直强度轴来定义的。HSI 彩色空间的重要分量是垂直强度轴、到一个彩色点的向量长度和该向量与红轴的夹角。因此，如图 6.13(c)和图 6.13(d)所示，HSI 平面以刚刚讨论的六边形、三角形，甚至一个圆形的形式出现也很常见。实际上选择什么形状无关紧要，因为这些形状中的任何一个都可以通过几何变换转换为其他两种形状。图6.14 显示了基于彩色三角形和圆形的 HSI 模型。

430 ~ 431

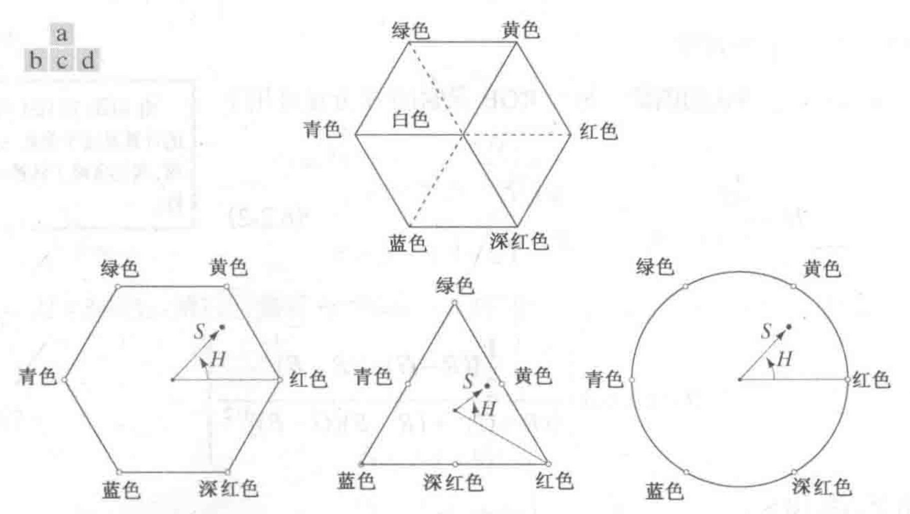


图 6.13　HSI 彩色模型中的色调和饱和度。黑点是一个任意彩色点。与红轴的夹角给出了色调，向量的长度是饱和度。这些平面中的所有彩色的强度由垂直强度轴上的平面的位置给出

图 6.14　基于(a)三角形和(b)圆形彩色平面的 HSI 彩色模型。三角形和圆形平面垂直于垂强度轴

431
~
432

从 RGB 到 HSI 的彩色转换

给定一幅 RGB 彩色格式的图像，每个 RGB 像素的 H 分量可用下式得到：

由 RGB 到 HSI 或从 HSI 到 RGB 的计算是逐个像素执行的。为表述清晰，我们省略了转换公式对 (x, y) 依赖性。

$$H = \begin{cases} \theta, & B \leqslant G \\ 360 - \theta, & B > G \end{cases} \tag{6.2-2}$$

式中[①]，

$$\theta = \arccos\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\}$$

饱和度分量由下式给出：

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \tag{6.2-3}$$

最后，强度分量由下式给出：

$$I = \frac{1}{3}(R+G+B) \tag{6.2-4}$$

如图 6.13 中指出的那样，假定 RGB 值已归一化到区间[0, 1]内，并且角度θ 根据 HSI 空间的红轴来度量。色调可以用式(6.2-2)得到的所有值除以 360° 归一化到区间[0, 1]内。如果给定的 RGB 值在区间[0, 1]内，则其他两个 HSI 分量也已在区间[0, 1]内。

式(6.2-2)到式(6.2-4)中的结果可由图 6.12 和图 6.13 所示的几何关系推得。这一推导很烦琐，而且对现在的讨论没有太大意义。对这些公式的证明及对下面的 HSI 至 RGB 转换结果感兴趣的读者，可以参考本书的参考文献或网站。

关于 RGB 和 HSI 之间及 HSI 和 RGB 之间的转换公式的详细推导过程，请读者参阅本书网站上的 Tutorials 部分。

从 HSI 到 RGB 的彩色转换

对于[0, 1]内的 HSI 值，现在我们想要在相同的值域找到对应的 RGB 值。可用的公式取决于 H 的值。在原色分隔中有 3 个相隔 120°的扇区(见图 6.13)。我们从 H 乘以 360°开始，这时的色调值回到原来的区间[0°, 360°]内。

RG 扇区($0° \leqslant H < 120°$)：当 H 的值在该扇区中时，RGB 分量由以下公式给出：

$$B = I(1-S) \tag{6.2-5}$$

$$R = I\left[1 + \frac{S\cos H}{\cos(60° - H)}\right] \tag{6.2-6}$$

$$G = 3I - (R+B) \tag{6.2-7}$$

GB 扇区($120° \leqslant H < 240°$)：如果给定的 H 值在该扇区中，则首先从 H 中减去 120°，即

$$H = H - 120° \tag{6.2-8}$$

① 在该表达式中加一个较小的数可避免在 $R = G = B$ 时被零除的情形(即θ 为 90°的情形)。注意，当所有的 RGB 分量相等时，由式(6.2-3)可得出 $S = 0$。此外，式(6.2-5)至式(6.2-7)中的由 HSI 至 RGB 的转换将给出 $R = G = B = I$，这是我们期望的结果，因为当 $R = G = B$ 时，我们处理的图像是灰度级图像。

432 ~ 433

则 RGB 分量为

$$R = I(1-S) \tag{6.2-9}$$

$$G = I\left[1+\frac{S\cos H}{\cos(60^\circ - H)}\right] \tag{6.2-10}$$

$$B = 3I-(R+G) \tag{6.2-11}$$

BR 扇区(240°≤ H < 360°)：最后，如果 H 的值在该扇区中，则从 H 中减去 240°，即

$$H = H-240^\circ \tag{6.2-12}$$

则 RGB 分量为

$$G = I(1-S) \tag{6.2-13}$$

$$B = I\left[1+\frac{S\cos H}{\cos(60^\circ - H)}\right] \tag{6.2-14}$$

$$R = 3I-(G+B) \tag{6.2-15}$$

这些公式在图像处理中的用法将在下面各节加以讨论。

例 6.2　对应于 RGB 彩色立方体图像的 HSI 值。

图6.15显示了图6.8中所示RGB值的色调、饱和度和强度图像。图6.15(a)是色调图像，其最大的特点是立方体前(红)平面沿 45°线的值是不连续的。为了解不连续的原因，可参见图6.8，从立方体的红色顶点到白色顶点画一条线，并在这条线的中间选择一点。从该点开始，向右环绕立方体一周直到回到开始点画一条轨迹。在该路径中出现的主要颜色是黄色、绿色、青色、蓝色、深红色、黑色直至红色。根据图 6.13，沿该路径的色调值应该从 0°增加到 360°（即从色调的最低可能值到最高可能值）。这正好如图 6.15(a)所示的那样，因为在灰度级中最低值代表黑色，最高值代表白色。事实上，这幅色调图像原本已被归一化到区间[0, 1]内，然后被缩放到 8 比特，即在显示时转换到了区间[0, 255]内。

在图6.15(b)中，饱和度图像显示了从暗值逐渐向 RGB 立方的白顶点的过渡，颜色越来越少，饱和度越来越低。最后，图6.15(c)所示强度图像中的每个像素值是图 6.8 中相应像素处 RGB 值的平均。

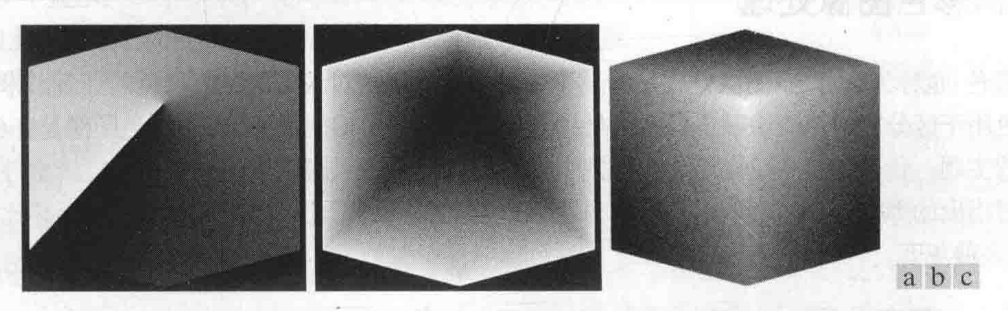


图 6.15　图 6.8 中图像的 HSI 分量：(a)色调图像；(b)饱和度图像；(c)强度图像

处理 HSI 分量图像

在下面的讨论中，我们考察处理 HSI 分量图像的几种简单技术，以便帮助你熟悉这些分量并加深对 HSI 彩色模型的理解。图 6.16(a)显示了一幅由 RGB 原色和二次色组成的图像。图 6.16(b)到图 6.16(d)显示了这幅图像的 H, S 和 I 分量图像，这些图像是用式(6.2-2)到式(6.2-4)产生的。回顾本节早些时候的讨论，图 6.16(b)中的灰度值对应于角度；例如，因为红色对应于 0°，所以图 6.16(a)

434～435

中的红区域被映射到色调图像中的一个黑色区域。类似地，图 6.16(c)中的灰度级对应于饱和度(为了显示，它们被放大到[0, 255])，图 6.16(d)中的灰度级是平均强度。

要在 RGB 图像中更改任何区域的个别颜色，我们可以改变图 6.16(b)所示色调图像中相应区域的值。然后用式(6.2-5)到式(6.2-15)中说明的步骤，将新的 H 图像和没有变化的 S 图像与 I 图像一起变换回 RGB 图像。要改变任何区域中颜色的饱和度(纯度)，除了在 HSI 空间改变饱和度之外，可采用相同的步骤。类似的方法可用于更改任何区域的平均强度。当然，这些改变可同时进行。例如，图 6.17(a)中的图像是通过将对应于图 6.16(b)中的蓝色、绿色区域的像素变为 0 得到的。在图 6.17(b)中，我们将图 6.16(c)的分量图像 S 中的青色区域的饱和度降低一半。在图6.17(c)中，我们把图 6.16(d)所示强度图像中心的白色区域的强度降低一半。把改进后的 HSI 图像转换回 RGB 彩色空间的结果显示在图6.17(d)中。如预期的那样，在这些图中我们看到，所有圆形的外部区域现在是红色的，青色区域的纯度降低了，中心区域变成了灰色而不再是白色。显然这些结果很简单，但却清楚地显示出了 HSI 模型独立控制色调、饱和度和强度的能力。当描述彩色时，我们会十分熟悉这些量。

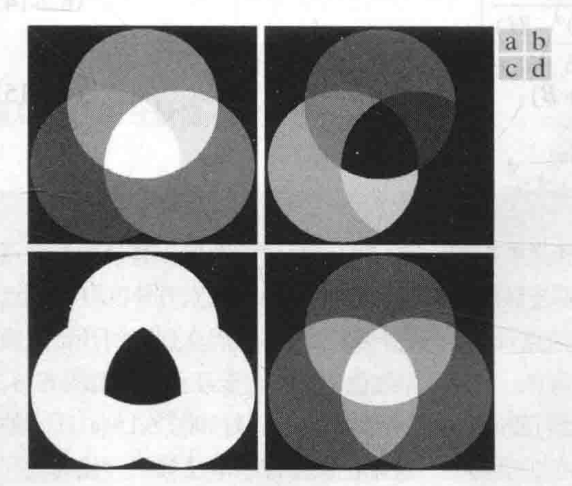


图 6.16 (a)RGB 图像及其对应的 HSI 图像的分量；(b)色调；(c)饱和度；(d)强度

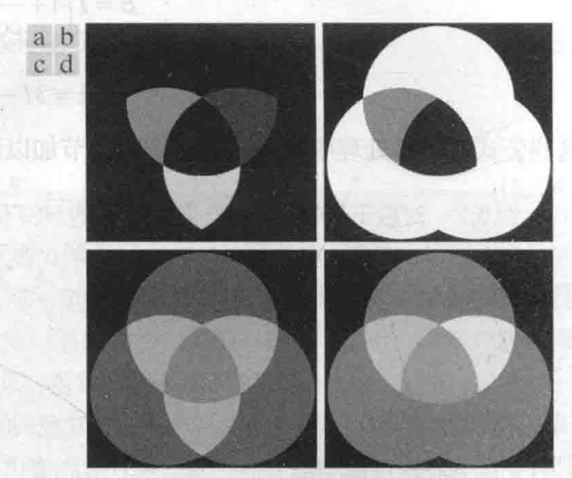


图 6.17 (a)~(c)修改后的 HSI 分量图像；(d)最终的RGB 图像(原始的 HSI 图像见图 6.16)

6.3 伪彩色图像处理

伪彩色(也称为*假彩色*)图像处理是指基于一种指定的规则对灰度值赋以颜色的处理。术语伪彩色或假彩色用于区分对单色图像赋予彩色的处理和与真彩色图像相关的处理，真彩色图像是在6.4节中开始讨论的主题。伪彩色的主要应用是人目视观察和解释单幅图像或序列图像中的灰度级事件。正如本章开始时指出的那样，利用彩色的主要动力之一是人类可以辨别几千种色调和强度，而相比之下只能辨别 20 多种灰度。

6.3.1 灰度分层①

灰度分层(有时称为*密度分层*)和彩色编码技术是伪彩色图像处理的最简单的例子之一。如果一幅图像被描述为三维函数［见图2.18(a)］，则分层方法可视为放置一些平行于该图像的坐标平面的平面，然后，每个平面在相交的区域中“切割”图像函数。图6.18显示了使用位于 $f(x,y)=l_i$ 处的一个平面把该图像函数切割为两部分的例子。

435
~
437

① 灰度分层在有些译著中也称为灰度分割——译者注。

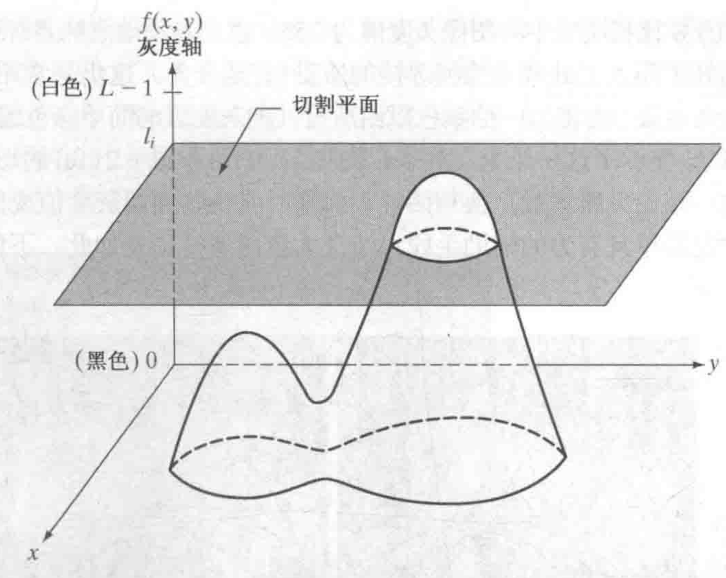


图 6.18　灰度分层技术的几何解释

如果对图6.18中平面的每一侧赋以不同的颜色，平面上面的任何灰度级的像素将编码成一种彩色，该平面之下的任何像素将编码成另一种颜色。位于平面上的灰度级本身被任意赋以两种彩色之一。结果是一幅只有两种颜色的图像，对应的外观可通过沿灰度轴上下移动这个切割平面来控制。

通常，该技术可总结如下：令$[0, L-1]$表示灰度级，令 l_0 代表黑色［$f(x, y)=0$］，并令 l_{L-1} 代表白色［$f(x, y)=L-1$］。假定垂直于灰度轴的 P 个平面定义为灰度级 $l_1, l_2, \cdots, l_P$。然后，假定 $0<P<L-1$，P 个平面将灰度分为 $P+1$ 个区间 $V_1, V_2, \cdots, V_{P+1}$。灰度级到彩色赋值根据如下关系进行：

$$f(x, y)=c_k \quad f(x, y) \in V_k \tag{6.3-1}$$

式中 c_k 是与第 k 个灰度区间 V_k 有关的颜色，V_k 由位于 $l=k-1$ 和 $l=k$ 处的分割平面定义。

平面的概念对于灰度分层技术的几何解释很有用。图6.19给出了另一种解释，它定义了如图 6.18 所示的相同映射。根据图 6.19 中显示的映射函数，任何输入的灰度级根据其是位于 l_i 值的上面还是下面而赋予两种颜色之一。当用更多的灰度级时，映射函数采用一种阶梯形式。

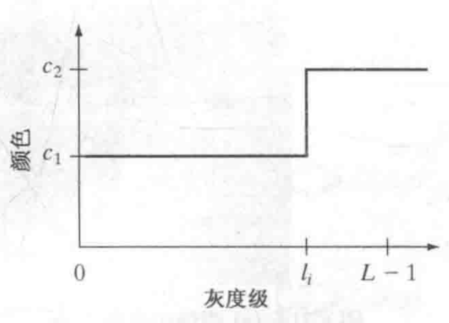


图 6.19　灰度分层技术的另一种表示

例 6.3　灰度分层。

一种简单而实用的灰度分层法示于图6.20。图6.20(a)是Picker甲状腺模型(放射实验模型)的单色图像，图 6.20(b)是灰度分层结果，图像分为 8 个彩色区域。单色图像中出现的恒定灰度区也完全可变，如切割后图像中的各种颜色所示。例如，左瓣在单色图像中是暗灰色的，因此以灰度形式分辨出病变很困难。相比之下，彩色图像清楚地显示出了恒定灰度区域的 8 个不同区域，其中每个区域采用一种颜色。

在前面这个简单的例子中，灰度被分为几个区间，且每个区域赋予了不同的颜色，而未考虑到图像中灰度级的意义。在这种情况下，感兴趣的只是简单地观察构成图像的不同灰度级。当灰度级细分基于图像的物理特性时，灰度分层就会有很大的意义并起重要作用。例如，图6.21(a)所示的含有裂纹和孔隙(水平通过图像中部的亮的白色条痕)的焊接物 X 光图像(水平暗区)。我们知道，当焊接物中有裂纹和孔隙时，X 射线的全部强度都会通过物体，使位于物体另一边的成像传感器饱和(见2.3节)。

437 ~ 439

这样，来自这一系统的 8 比特图像中的图像灰度值为 255，进而自动地意味着焊接有问题。如果人是分析的最终判断者，并采用人工处理来检测焊接的质量(直至今天，这仍是常用的工序)，那么将一种颜色赋给 255 这个灰度级、而将另一种颜色赋给所有其他灰度级的简单彩色编码，会极大地简化检测者的工作。图 6.21(b)显示了这一结果。如果以图 6.21(b)而非图 6.21(a)的形式来显示图像，“人的误识率将降低”这一结论毋庸置疑。换句话说，如果所寻找的确切灰度值或值域是已知的，则灰度分层在可视化方面是简单且有力的辅助手段，涉及大量图像时更是如此。下面是一个更为复杂的例子。

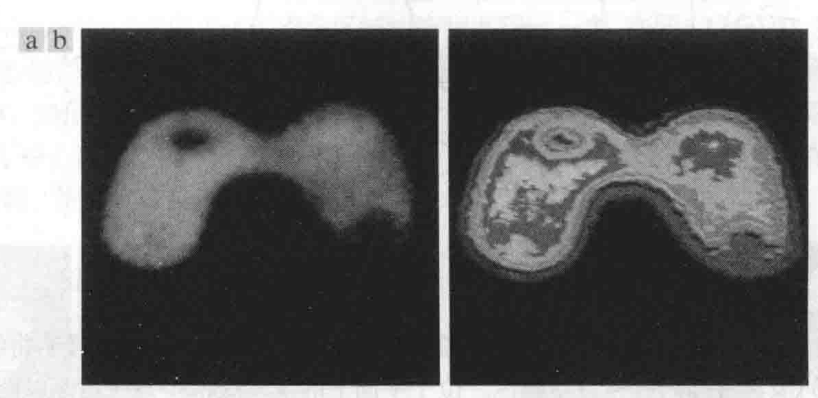


图 6.20 (a) Picker 甲状腺模型的单色图像；(b) 密度分层为 8 种颜色的结果(原图像由美国 Oak Ridge 国家实验室仪器与控制分部的 J. L. Blankenship 博士提供)

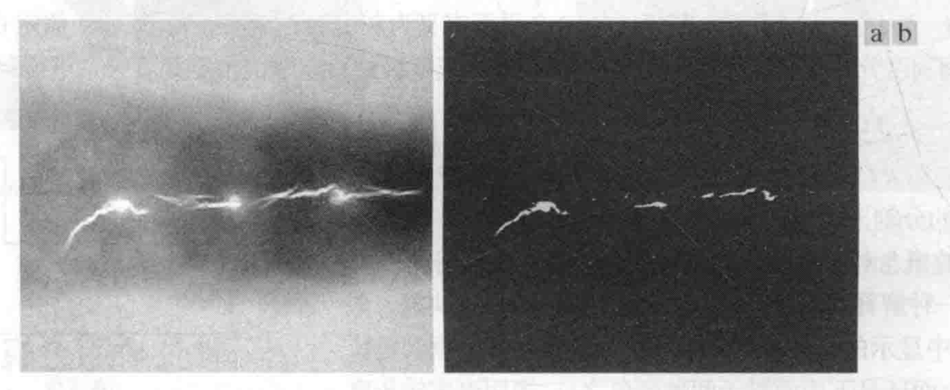


图 6.21 (a) 焊缝的单色 X 射线图像；(b) 彩色编码的结果(原图像由 X-TEK 系统有限公司提供)

例 6.4 使用彩色突出降雨量。

测量地球上各区域(尤其是热带区域)的降雨量，在与环境有关的各种应用中很重要。用地面传感器准确测量降雨量很困难，而且代价高昂，得到总雨量图更为困难，因为重要的降雨区都在海洋上。获得降雨量图的一种方法是使用卫星。TRMM(Tropical Rainfall Measuring Mission，热带降雨量测量任务)卫星使用为检测降雨而特殊设计的 3 个传感器：一个降雨雷达、一个微波成像仪和一个可见光及红外扫描仪(关于图像传感模式，请参阅 1.3 节和 2.3 节)。

来自各个雨量传感器的结果经处理后，得出传感器监视区域在给定时间段的平均降雨量的估计。从这些估计生成灰度值直接对应于降雨量的灰度图像并不困难，每个像素代表自然的陆地面积，其大小取决于传感器的分辨率。这样的一幅灰度图像示于图6.22(a)，其中，由卫星监视的区域是在图像中间的三分之一处稍亮一些的水平带(即热带区域)。在这个特殊的例子中，降雨值是三年时间内的月平均值(单位为英寸)。

442 ~ 444

对于降雨模式，目测这些图像也是十分困难的，甚至是不可能的。然而，假设使用图6.22(b)所示的彩色对 0 到 255 的灰度级编码，那么趋于蓝色的值意味着低降雨量，相反则为红色。注意，标度终止于纯红色处，表示降雨量大于 20 英寸。图6.22(c)显示了使用刚才讨论的彩色映射对灰度图像进行彩色编码的结果。如图 6.22(c)和图6.22(d)的放大区域所示，这些结果更易于解释。另外，除了提供全球覆盖外，这种类型的数据允许气象工作者使用比以往更精确的方法来标定地面的雨量监控系统。

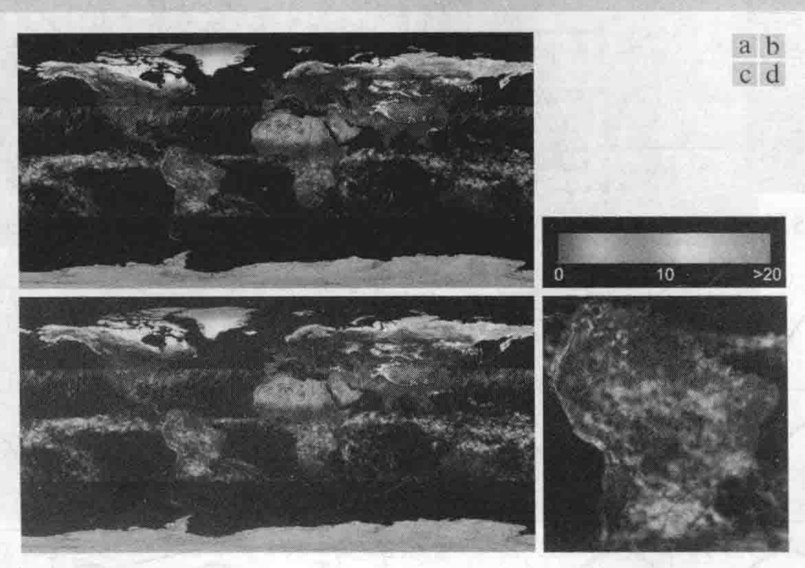


图 6.22　(a) 与月平均降雨相对应的灰度图像；(显示为较亮的水平条带) (b) 赋给灰度值的颜色；(c) 彩色编码的图像；(d) 南美区域的放大图(NASA 提供)

6.3.2　灰度到彩色的变换

与前一节讨论的简单分层技术相比，其他类型的变换更通用,也更能拓宽伪彩色增强结果的范围。一种特别有吸引力的方法如图6.23 所示。这一方法的基本概念是，对任何输入像素的灰度执行 3 个独立的变换。然后，将 3 个变换结果分别送入彩色电视监视器的红、绿、蓝通道。这种方法产生一幅合成图像，合成图像的彩色内容由变换函数的特性调制。注意，这些方法只对一幅图像的灰度值进行变换，而变换位置。

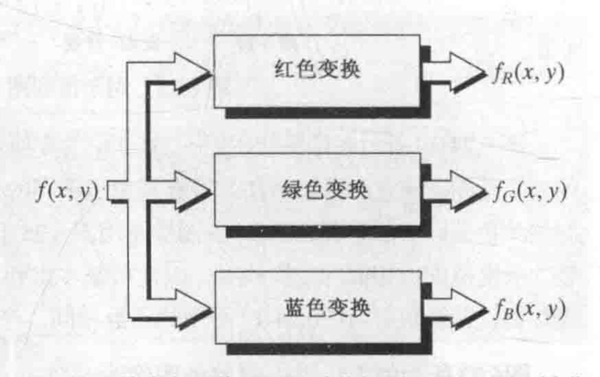


图 6.23　伪彩色图像处理的功能框图。f_R、f_G和f_B被分别送入 RGB 彩色监视器的红、绿和蓝输入

前一节讨论的方法是刚描述的技术的一种特殊情况。这里，灰度级的分段线性函数(见图 6.19)用来产生颜色。另一方面，本节讨论的方法是以平滑的非线性函数为基础的，因此这就在技术上给出了相当大的灵活性。

例 6.5　突出行李内包含的爆炸物的伪彩色应用。

图 6.24(a)显示了由机场 X 射线扫描系统得到的行李的两幅单色图像。左侧的图像含有普通物品，右侧的图像含有相同的物品及一块塑料仿真爆炸物。这个例子的目的是说明使用灰度级到彩色的变换可得到不同程度的增强。

图6.25显示了所用的变换函数。这些正弦函数包含有在波峰附近相对恒定的区域，以及在波谷附近迅速变化的区域。改变每个正弦函数的相位和频率可(以彩色)强调灰度级范围。例如，如果所有 3 个变换都

440 ~ 442

有相同的相位和频率，则输出图像将是单色的。3 个变换函数间相位上的细小变化可产生对应于正弦波峰的灰度像素的微小变化，特别是在正弦波有着较宽的曲线(低频)时。由于正弦波间的相移导致的 3 个正弦波的幅度明显不同，位于正弦波陡峭部分的那些灰度级的像素被赋予了更强的彩色内容。

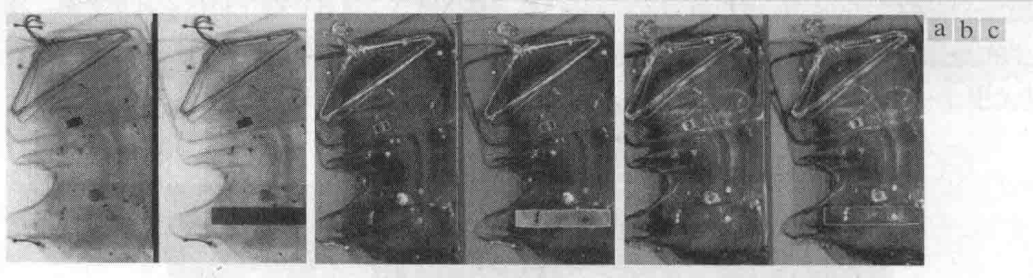


图 6.24　使用图 6.25 所示灰度级到彩色的变换的伪彩色增强(原图像由 Westinghouse 公司的 Mike Hurwitz 博士提供)

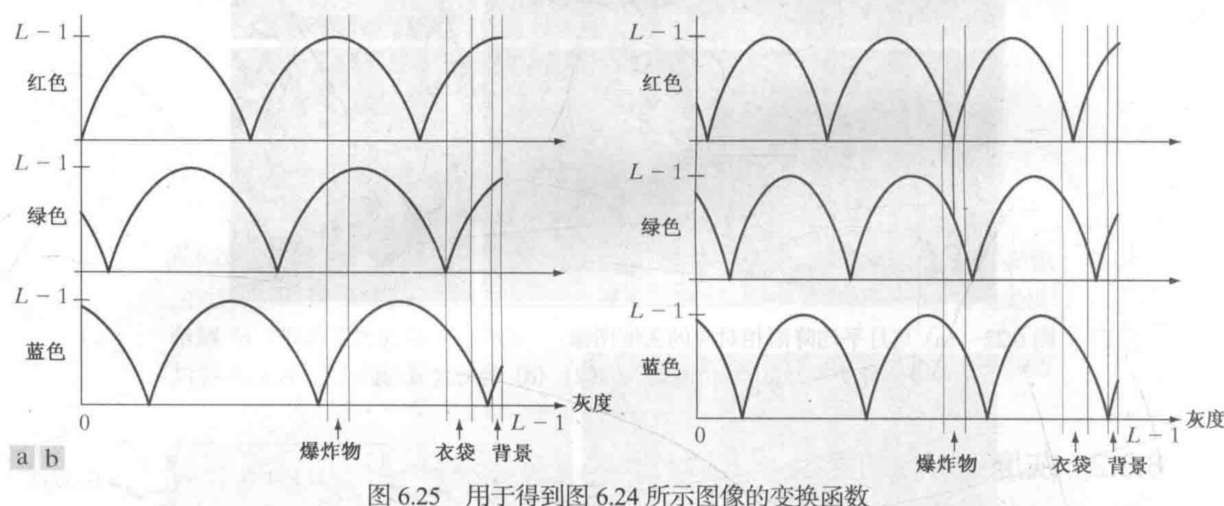


图 6.25　用于得到图 6.24 所示图像的变换函数

图 6.24(b)所示图像是使用图 6.25(a)中的变换函数得到的，它显示了与爆炸物、衣袋和背景分别对应的灰度范围。注意，爆炸物和背景有着相当不同的灰度级，但是由于正弦波的周期性，它们都使用近似相同的颜色编码。示于图 6.24(c)的图像是用图 6.25(b)所示的变换函数得到的。在这种情况下，爆炸物和衣袋的灰度范围由相似的变换映射，因此可基本得到相同的彩色赋值。注意，这种映射允许观察者“看透”爆炸物。背景映射与图6.24(b)所用的函数相同，产生几乎相同的彩色赋值。

图6.23所示的方法以一幅单色图像为基础。通常，如图6.26所示，将多幅单色图像组合为一幅彩色合成图像是很重要的。在多光谱图像处理中常使用这一方法(如例 6.6 所示)，这里，不同的传感器在不同的谱段产生独立的单色图像。图 6.26 所示的附加处理可以是彩色平衡(见 6.5.4 节)、混合图像，以及以产生图像的传感器响应特性的知识为基础来选择三幅用于显示的图像等技术。

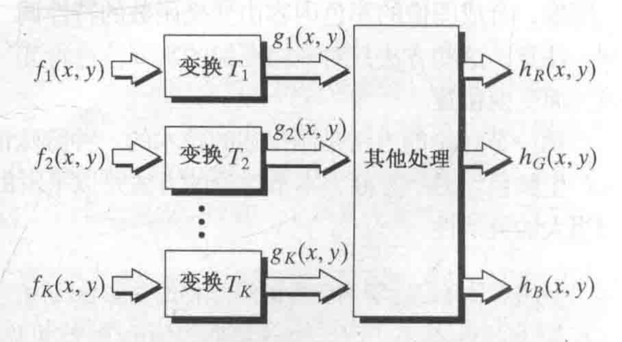


图 6.26　几幅单色图像可用时，所用的伪彩色编码方法

例 6.6　多光谱图像的彩色编码。

442 ~ 444

图6.27(a)到图6.27(d)显示了 4 幅华盛顿特区的光谱卫星图像，包括波托马克河的一部分。前 3 幅图像是可见光的红、绿、蓝图像，第 4 幅图像是近红外图像(见表1.1 和图1.10)。图6.27(e)是把前 3 幅图像

合成为一幅 RGB 图像后得到的全彩色图像。密集区域的全彩色图像很难判读，但该图像的一个显著特点是波托马克河的各部分颜色不同。图 6.27(f) 较为有趣，这幅图像是用近红外图像代替图 6.27(e) 的红色分量形成的。从表 1.1 可知，这一波段对场景中的生物分量有较强的反应。图6.27(f) 显示了生物(红色) 和场景中人为特征间的明显区别，由混凝土和柏油组成的部分在图像中显示为浅蓝色。

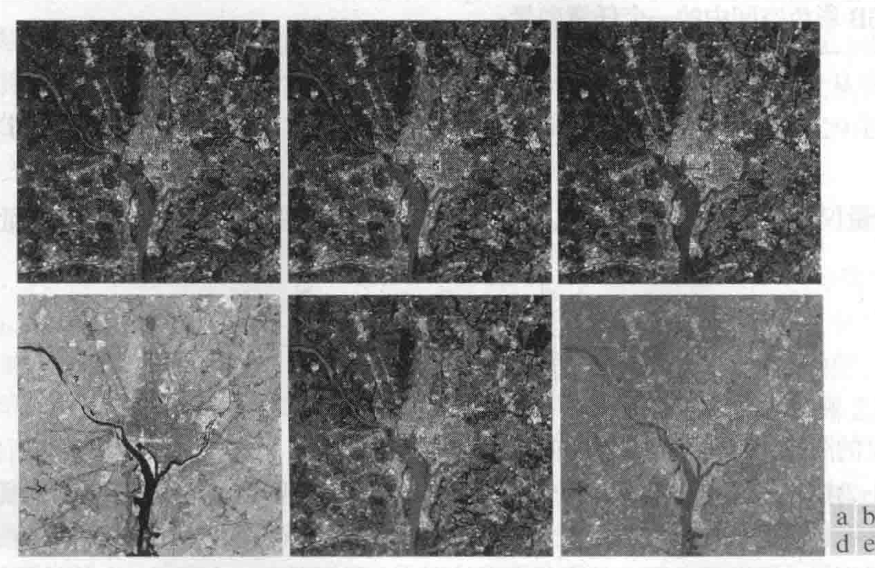


图 6.27　(a)~(d) 图 1.10 中的 1 到 4 波段的图像(见表 1.1)；(e) 将图(a)、(b) 和(c) 作为 RGB 图像的红、绿、蓝分量得到的彩色合成图像；(f) 用相同的方法，但在红通道使用(d) 中的近红外图像得到的图像(原始的多光谱图像由 NASA 提供)

刚才说明的处理类型，在复杂图像中对感兴趣物体进行可视化处理的功能很强大，特别是当这些物体超出了我们的正常感知能力时。图 6.28 很好地说明了这一点。这些图像是通过把来自“伽利略”宇宙飞船的几幅传感器图像组合在一起并以伪彩色方式显示的木卫一的图像，其中一些位于人眼不可见的光谱范围。然而，通过了解影响传感器响应的物理和化学处理，将感知图像组合为一幅有意义的伪彩色图中是可能的。组合感知图像数据的一种方法在于如何显示表面化学成分的差别，或改变表面反射阳光的方式。例如，在图 6.28(b) 所示的伪彩色图像中，亮红色表示木卫一上上活火山最近喷出的物质，而周围的黄色物质是陈旧的硫沉积物。这幅图像传达这些特征要比分别分析分量图像更容易。

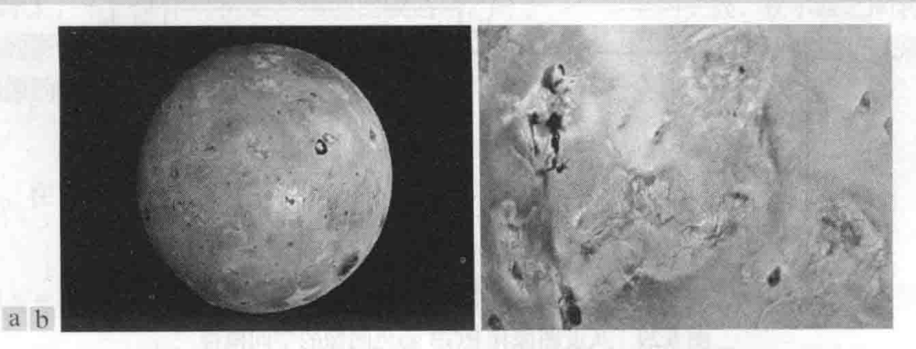


图 6.28　(a) 木卫一的伪彩色图像；(b) 靠近摄取的图像(原图像由 NASA 提供)

6.4　全彩色图像处理基础

本节介绍适用于全彩色图像的处理技术。虽然这种技术的研究远未穷尽，但本节接下来阐述的技术说明了面对各种图像处理任务时如何处理彩色图像。全彩色图像处理方法分为两大类。第


444
~
446

一类是分别处理每一幅分量图像，然后由分别处理过的分量图像来形成一幅处理过的合成彩色图像。第二类是直接处理彩色像素。因为全彩色图像至少有 3 个分量，所以彩色像素实际上是向量。例如，在 RGB 系统中，每个彩色点可以用 RGB 坐标系统中从原点延伸到该点的一个向量来解释(见图 6.7)。

令 $\mathbf{c}$ 代表 RGB 彩色空间中的一个任意向量：

$$\mathbf{c} = \begin{bmatrix} c_R \\ c_G \\ c_B \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{6.4-1}$$

该式指出，$\mathbf{c}$ 的分量仅是一幅彩色图像在一点处的 RGB 分量。我们可以这样认为，彩色分量是坐标(x, y)的函数，表示为

$$\mathbf{c}(x,y) = \begin{bmatrix} c_R(x,y) \\ c_G(x,y) \\ c_B(x,y) \end{bmatrix} = \begin{bmatrix} R(x,y) \\ G(x,y) \\ B(x,y) \end{bmatrix} \tag{6.4-2}$$

对于大小为 $M \times N$ 的图像，有 MN 个这样的向量 $\mathbf{c}(x,y)$，其中 $x = 0, 1, 2, \cdots, M-1$； $y = 0, 1, 2, \cdots, N-1$。

记住，式(6.4-2)描述的向量很重要，其分量是关于 x 和 y 的空间变量。这常常是产生混淆的根源，但通过将重点放在我们感兴趣的空间处理上，可以避免混淆。也就是说，我们对以 x 和 y 的公式来表达图像处理技术感兴趣。根据像素现在是彩色像素的事实，我们引入一个因子，在其最简单的公式中，允许我们用标准的灰度图像处理方法分别处理彩色图像的每个分量。然而，单个彩色分量处理的结果并不总是等同于彩色向量空间中的直接处理，在这种情况下，我们必须采用新的表达方法。

每种彩色分量的处理要等同于基于向量的处理，必须满足两个条件：第一，处理必须对向量和标量都可用；第二，对向量的每个分量的操作对于其他分量必须是独立的。作为说明，图6.29 显示了灰度图像和全彩色图像的邻域空间处理。假定该处理是邻域平均。在图6.29(a)中，平均是将邻域内的所有像素的灰度相加，然后除以邻域内的像素总数。在图6.29(b)中，平均是把邻域内的所有向量相加，并用邻域内的向量总数去除每个分量。但平均向量的每个分量是对应于该分量的图像中的像素总和，这与在每个彩色分量基础上做平均，然后再形成向量所得到的结果是相同的。接下来的几节将详细说明这一处理。此外，还将给出一种方法，采用这种方法时，两种处理结果是不同的。

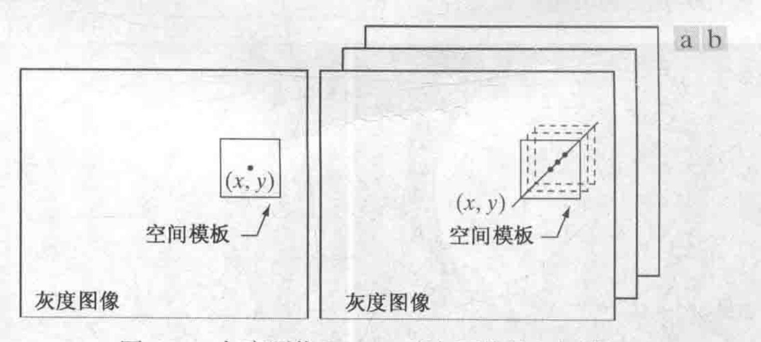


图 6.29 灰度图像和 RGB 彩色图像的空间模板

6.5 彩色变换

447 ~ 448

本节讨论的技术总称为彩色变换，主要涉及在单一彩色模型内处理彩色图像的分量，而不是这些分量在不同模型间的转换(如 6.2.3 节的 RGB 到 HSI 和 HSI 到 RGB 的转换)。

6.5.1　公式

如第 3 章的灰度变换技术那样，我们用下式表达的彩色变换作为模型：

$$g(x,y)=T\left[f(x,y)\right] \tag{6.5-1}$$

式中 $f(x, y)$是彩色输入图像，$g(x, y)$是变换后或处理过的彩色输出图像，T 是在(x, y)的空间邻域上对 f 操作的一个算子。该式与式(3.1-1)的主要区别在于其描述不同。这里的像素值是从彩色空间选择的、用来表示图像的彩色空间的 3 元组或 4 元组(即 3 个值或 4 个值的组合)，如图6.29(b)所示。

类似于 3.2 节中用于介绍基本灰度变换的方法，本节只关注形如

$$s_i=T_i(r_1,r_2,\cdots,r_n),\quad i=1,2,\cdots,n \tag{6.5-2}$$

的彩色变换，其中，为标记简单起见，r_i 和 s_i 是 $f(x, y)$和 $g(x, y)$在任何点处彩色分量的变量，n 是彩色分量数，$\{T_1, T_2,\cdots, T_n\}$是对 r_i 操作产生 s_i 的一组变换或彩色映射函数。注意，n 个变换 T_i 合并可执行式(6.5-1)中的单一变换函数 T。用于描述 f 和 g 的像素的所选彩色空间决定 n 的值。例如，如果选择 RGB 彩色空间，则 $n=3$，且 r_1, r_2 和 r_3 分别表示输入图像的红、绿、蓝分量。如果选择 CMYK 或 HSI 彩色空间，则 $n=4$ 或 $n=3$。

图6.30显示了一碗草莓和一杯咖啡的高分辨率彩色图像，它是对大幅面(4 英寸 × 5 英寸)彩色底片数字化后得到的图像。图中的第二行包含原始的 CMYK 扫描分量图像。在这些图像的每个 CMYK 彩色分量中，黑用 0 表示，白用 1 表示。这样，我们看到草莓由大量的深红色和黄色组成，因为对应于这两种 CMYK 分量的图像最亮。黑色较少，且通常仅出现在咖啡和草莓碗的阴影中。当 CMYK 图像被转换为 RGB 时，如图中第三行所示，可以看到草莓包含大量的红色和很少的绿色与蓝色。图 6.30 的最后一行显示了使用式(6.2-2)到式(6.2-4)计算出的图6.30(a)的 HSI 分量图像。如所期望的那样，强度分量是全彩色原始图像的单色再现。另外，草莓的颜色相对较为纯净。它们具有最高的饱和度，或图像中色调被白光稀释得最少。最后，我们注意到解释这幅色度分量图像时的一些困难。其中包含这样一些事实：(1)在 HSI 模型中，0°和 360°相遇处有一个不连续点；(2)色调对于 0 饱和度(即对白色、黑色和纯灰色)未定义。模型的不连续点多出现在草莓周围，它们用接近白(1)和黑(0)的灰度值描述，结果是用不希望的高对比灰度级来表示单一颜色——红色。

图6.30中的任何彩色空间分量都可与式(6.5-2)共同运用。理论上，任何变换都可在任何彩色模型中执行。但实际上，某些操作对特定的模型比较适用。对于一个给定的变换，表示法之间转换的代价是必须判定相关的彩色空间，然后再在该空间上执行该转换。例如，假设我们想要使用下式来改进图 6.30 中的全彩色图像的强度：

$$g(x,y)=kf(x,y) \tag{6.5-3}$$

式中 $0<k<1$。在 HSI 彩色空间中，使用如下的简单变换来完成：

$$s_3=kr_3 \tag{6.5-4}$$

式中 $s_1=r_1$ 和 $s_2=r_2$。仅有 HSI 强度分量 r_3 被改变。在 RGB 彩色空间中，3 个分量都必须变换：

$$s_i=kr_i,\quad i=1,2,3 \tag{6.5-5}$$

CMY 空间要求一组类似的线性变换集：

$$s_i=kr_i+(1-k),\quad i=1,2,3 \tag{6.5-6}$$

虽然 HSI 变换包含的操作最少，但把 RGB 图像或 CMYK 图像转换至 HSI 空间所要求的计算量大大抵消了简单变换的优点，即转换的计算量比强度变换本身的计算量大。然而，若不考虑所选择

448 ~ 450

的彩色空间，则其输出是相同的。图 6.31(b)显示了采用式(6.5-4)到式(6.5-6)对图 6.30 所示全彩色图像用 $k = 0.7$ 进行变换后的结果。映射函数示于图6.31(c)到图 6.31(e)中。

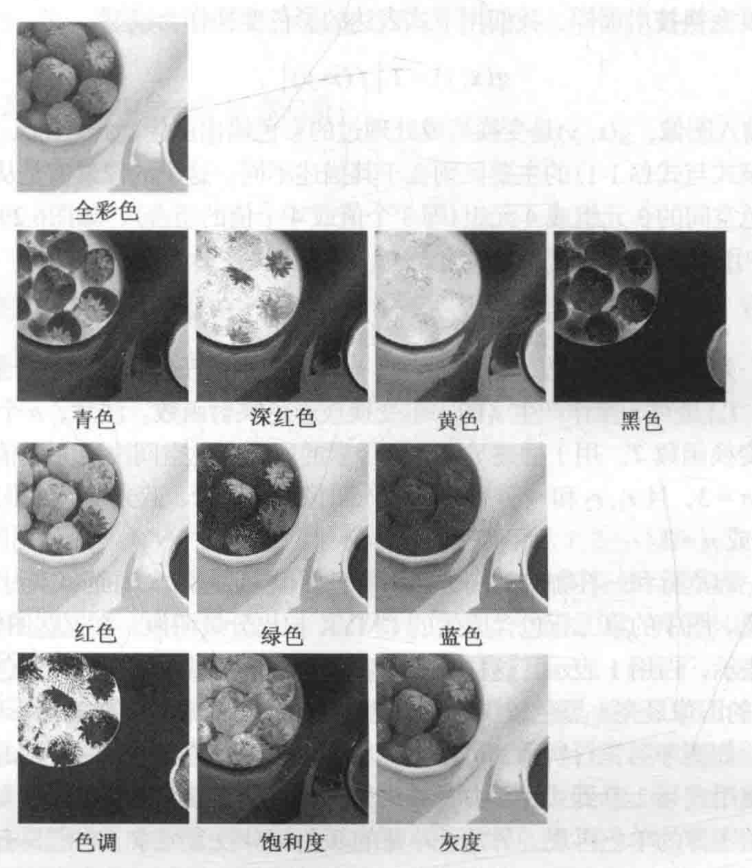


图 6.30 全彩色图像及其各种彩色空间分量(原图像由 MedData Interactive 公司提供)

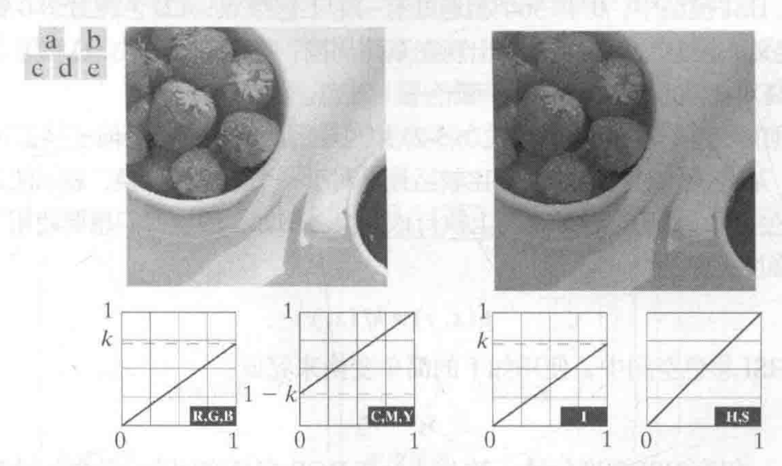


图 6.31 用彩色变换调整一幅图像的亮度：(a)原图像；(b)亮度降低 30%后的结果(即 $k = 0.7$)；(c)~(e)所需的 RGB, CMY 和 HSI 变换函数(原图像由 MedData Interactive 公司提供)

注意，式(6.5-4)到式(6.5-6)定义的每个变换，只与彩色空间的一个分量相关，这一点很重要。例如，红输出分量 S_1 在式(6.5-5)中独立于绿(r_2)输入和蓝(r_3)输入，它只依赖于红(r_1)输入。正如开始讨论时所提到的那样，这类变换是最简单和最常用的彩色处理工具，并且可以在每个彩色分量上执行。在本节余下的部分，我们研究几种这样的变换并讨论这样一种情况，在这种情况下，分量变换

450 ~ 451

函数依赖于输入图像的所有彩色分量，因此不能以单独的彩色分量为基础进行变换。

6.5.2　补色

在图 6.32 所示的彩色环[①]上，与色调直接相对的另一端被称为补色。对补色的兴趣源于这样一个事实，即补色类似于 3.2.1 节的灰度负值。正如在灰度情况下那样，补色对于增强嵌在彩色图像暗区的细节很有用——特别是区域在大小上占优势时。

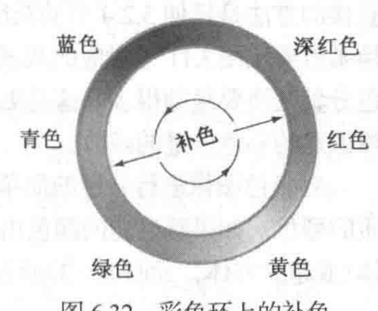


图 6.32　彩色环上的补色

例 6.7　计算彩色图像的补色。

图 6.33(a) 和图 6.33(c) 显示了来自图 6.30 的全彩色图像及其补色图像。用于计算补色的 RGB 变换如图 6.33(b) 所示。它们等同于 3.2.1 节中定义的灰度负变换。注意，计算的补色会使人联想到普通照片的彩色底片。原图像的红色被补色中的青色替代。当原图像是黑色时，补色是白色，以此类推。补色图像中每种色调都用图 6.32 的彩色环由原图像预测，且涉及补色计算的每个 RGB 分量变换只是对应的输入彩色分量的函数。

与图6.31 的亮度变换不同，在这个例子中使用的 RGB 补色变换函数并没有 HSI 空间的直接等效量。我们将其作为一个练习留给读者(见习题6.18)，要求说明补色的饱和度分量不能单独从输入图像的饱和度分量计算出来。图6.33(d) 提供了一种使用图 6.33(b) 给出的色调、饱和度和亮度变换的补色的近似。注意，输入图像的饱和度分量不可改变；它造成了图 6.33(c) 和图 6.33(d) 之间的视觉差别。

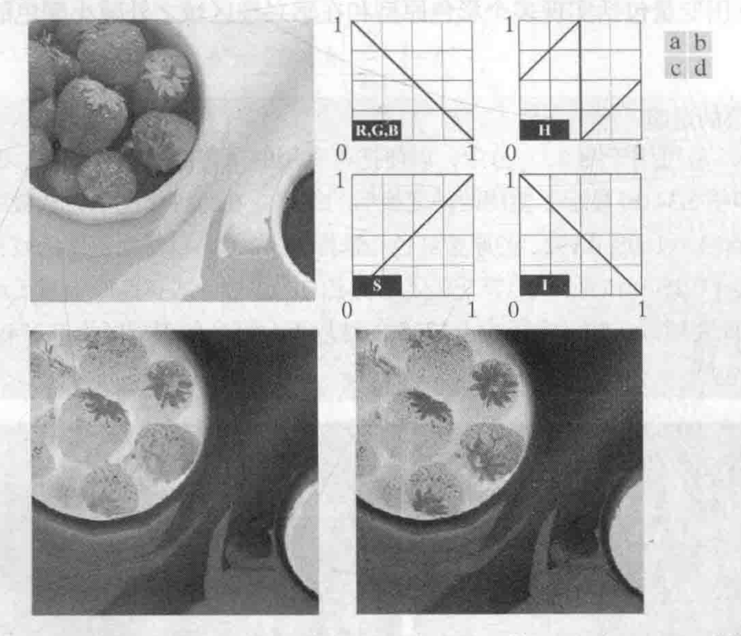


图 6.33　补色变换：(a) 原图像；(b) 补色变换函数；(c) 基于 RGB 映射函数的图 (a) 的补色；(d) 使用 HSI 变换的 RGB 补色的近似

6.5.3　彩色分层

突出图像中某个特定彩色区域对从其周围分离出目标物体很有用。其基本思路不是 (1) 显示感兴趣的颜色以便从背景中突出它们，就是 (2) 像模板那样使用由彩色定义的区域，以便进一步处理。最

451
~
453

① 该彩色环源自艾萨克·牛顿 (Isaac Newton) 先生，他于 17 世纪连接了彩色谱的末端从而形成了第一个彩色环。

直接的方法是延伸 3.2.4 节的灰度分层技术。然而，因为一个彩色像素是一个 n 维参量，得到的彩色变换函数要比图3.11 中对应的灰度变换函数复杂得多。事实上，所要求的变换比到目前为止所考虑的彩色分量变换要复杂得多。这是因为所有彩色分层方法都要求每个像素变换后的彩色分量是所有 n 个原始像素的彩色分量的函数。

对彩色图像进行分层的简单方法之一是，把某些感兴趣区域之外的彩色映射为不突出的无确定性质的颜色。如果感兴趣的颜色由宽为 W、中心在原型(即平均)颜色点并具有分量$(a_1, a_2, \cdots, a_n)$的立方体(或超立方体，此时 $n > 3$)所包围，则必要的一组变换为

$$s_i = \begin{cases} 0.5, & \left[\left|r_j - a_j\right| > \dfrac{W}{2}\right]_{1\le j\le n} \\ r_i, & \text{其他} \end{cases} \quad i = 1, 2, \cdots, n \tag{6.5-7}$$

这些变换通过强迫所有其他颜色为参考彩色空间的中点(一个任意选取的中性点)，来突出原型周围的颜色。例如，对于 RGB 彩色空间，合适的中性点是灰度或彩色的中点(0.5, 0.5, 0.5)。

如果使用一个圆球体来指定感兴趣的颜色，则式(6.5-7)变为

$$s_i = \begin{cases} 0.5, & \sum_{j=1}^{n}(r_j - a_j)^2 > R_0^2 \\ r_i, & \text{其他} \end{cases} \quad i = 1, 2, \cdots, n \tag{6.5-8}$$

式中 R_0 是封闭球形(或超球形，此时 $n > 3$)的半径，$(a_1, a_2, \cdots, a_n)$是其中心的分量(即原型颜色)。式(6.5-7)和式(6.5-8)的其他有用变量包括实现多个彩色原型和在感兴趣区域之外减小颜色的亮度，而不是对它们赋予中性值。

例 6.8　彩色分层的说明。

式(6.5-7)和式(6.5-8)可用于图 6.31(a)中，以便把可食用的草莓部分从背景杯、碗、咖啡和桌子中分离出来。图 6.34(a)和图 6.34(b)显示了使用两种变换后的结果。在每种情况下，从最突出的草莓中选择具有 RGB 彩色坐标(0.6863, 0.1608, 0.1922)的原型红色；选择 W 和 R_0，以便使重要区域不会扩展到不希望的图像区域。实际值 $W = 0.2549$ 和 $R_0 = 0.1765$ 是交互确定的。注意，式(6.5-8)中基于球形的变换在包含更多草莓的红色区域方面要稍好一些。半径为 0.1765 的球形不会完全包围宽度为 0.2549 的立方体，但它本身也不被立方体完全包围。

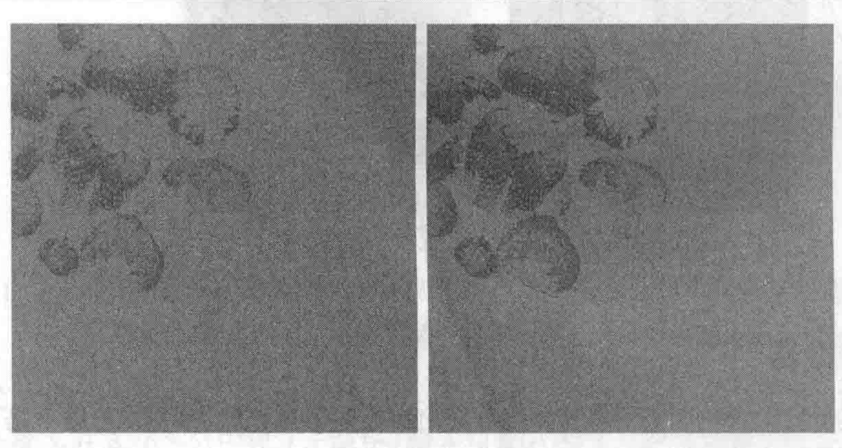


图 6.34　(a)检测宽度为 $W = 0.2549$、中心在(0.6863, 0.1608, 0.1922)的 RGB 立方体内的红色的彩色分层变换；(b)检测半径为 0.1765、中心在相同点的 RGB 球形中的红色的彩色分层变换。处在立方体和球形之外的像素由颜色(0.5, 0.5, 0.5)代替

453 ~ 455

6.5.4　色调和彩色校正

彩色变换可在多数台式计算机上执行。与数字摄像机、平板扫描仪和喷墨打印机相连，个人计算机就变成了数字暗室，从而允许我们对图像进行色调调整和彩色校正；高级彩色重现系统不需要配备传统的湿式处理(暗室)设备即可进行这些处理。虽然色调和彩色校正在其他成像领域也很有用，但当前讨论重点关注的是其最普通的应用——照片增强和颜色再现。

本节研究的变换效果最终要在打印效果中评价。因为这些变换是在监视器上研究、改进和评价的，所以在所用监视器和最终输出设备之间保持色彩的高度一致性是必要的。事实上，监视器的颜色应准确地表示任何数字扫描的源图像和最后的打印输出。最好使用独立于设备的彩色模型来实现，而彩色模型与监视器、输出设备及所用其他设备的色域有关(见 6.1 节)。这种方法把每个设备映射到模型及模型本身的色彩分布质量函数。许多彩色处理系统(CMS)选择的模型是 CIE $L^*a^*b^*$模型，也称为 CIELAB(CIE[1978], Robertson[1977])。$L^*a^*b^*$彩色分量由如下公式给出：

$$L^* = 116 \cdot h\left(\frac{Y}{Y_W}\right) - 16 \tag{6.5-9}$$

$$a^* = 500\left[h\left(\frac{X}{X_W}\right) - h\left(\frac{Y}{Y_W}\right)\right] \tag{6.5-10}$$

$$b^* = 200\left[h\left(\frac{Y}{Y_W}\right) - h\left(\frac{Z}{Z_W}\right)\right] \tag{6.5-11}$$

式中，

$$h(q) = \begin{cases} \sqrt[3]{q}, & q > 0.008\,856 \\ 7.787q + 16/116, & q \leqslant 0.008\,856 \end{cases} \tag{6.5-12}$$

且 X_W, Y_W 和 Z_W 是参考白色三激励值，即 CIE 标准 D65 照度下的一种完美的漫反射白色(在图 6.5 所示的 CIE 色度图中，由 $x = 0.3127$ 和 $y = 0.3290$ 定义)。$L^*a^*b^*$彩色空间是比色的(即与感觉色相匹配的编码相同)、感觉一致的(即在各种色调中颜色差别感觉上是一致的，见 Mac Adams[1942]的经典论文)和独立于设备的。虽然没有可直接显示的格式(需要变换到其他彩色空间)，但其色域包括整个可见光谱，并可以准确地表示任何显示设备、打印设备或输入设备的颜色。类似 HSI 系统，$L^*a^*b^*$系统是一种优秀的亮度(由亮度 L^*表示)和彩色(a^*表示红减绿，b^*表示绿减蓝)分离器，它在图像操作(色调和对比度编辑)和图像压缩方面很有用①。

标定成像系统的主要意义是，它允许交互地和独立地校正色调与色彩的不平衡，即按两个顺序操作。在解决诸如过饱和及欠饱和这样的不规则色彩问题之前，先要校正图像的色调范围。一幅图像的色调范围(也称为*主调类型*)是指颜色强度的基本分布。高主调图像的多数信息集中在高(亮)强度处；低主调图像的颜色主要位于低亮度处；中主调图像的颜色则位于两者之间。如单色情况下那样，通常期望一幅彩色图像的强度均匀分布在高亮和阴影之间。下面的几个例子说明了用于校正色调和色彩不平衡的各种彩色变换。

① 研究表明，$L^*a^*b^*$中照度(亮度)信息与彩色信息分离的程度要大于其他彩色模型中［如 CIELUV, YIQ, YUV, YCC 和 XYZ(Kasson and Plouffe[1992])］的相应程度。

455 ~ 456

例 6.9　色调变换。

改善图像色调的变换通常需要交互选择。想法是试验性地调整图像亮度和对比度，以便在合适的灰度范围提供最多的细节。彩色本身并不改变。在 RGB 和 CMY(K)空间中，这意味着使用相同的变换函数映射所有 3(或 4)个彩色分量；在 HSI 彩色空间中，则改进了亮度分量。

图6.35显示了用于校正 3 种常见色调不平衡的几个典型变换——即平淡的、较亮的和较暗的图像。图中第一行的 S 形曲线对提升对比度较为理想［见图3.2(a)］。其中的点被固定，因此高亮区域和阴影区域可分别变暗和加亮(该曲线反过来可用于校正过大的对比度)。图中第二行和第三行中的变换校正较亮图像和较暗图像，并使人联想到图 3.6 中的幂律变换。虽然彩色分量是离散的，如现在的实际变换函数，但这些变换函数本身以连续量显示和操作——通常由分段线性或高阶(对平滑映射)多项式构造。注意，在图6.35 中，图像的主调可直接观察；它们还可以用图像彩色分量的直方图来确定。

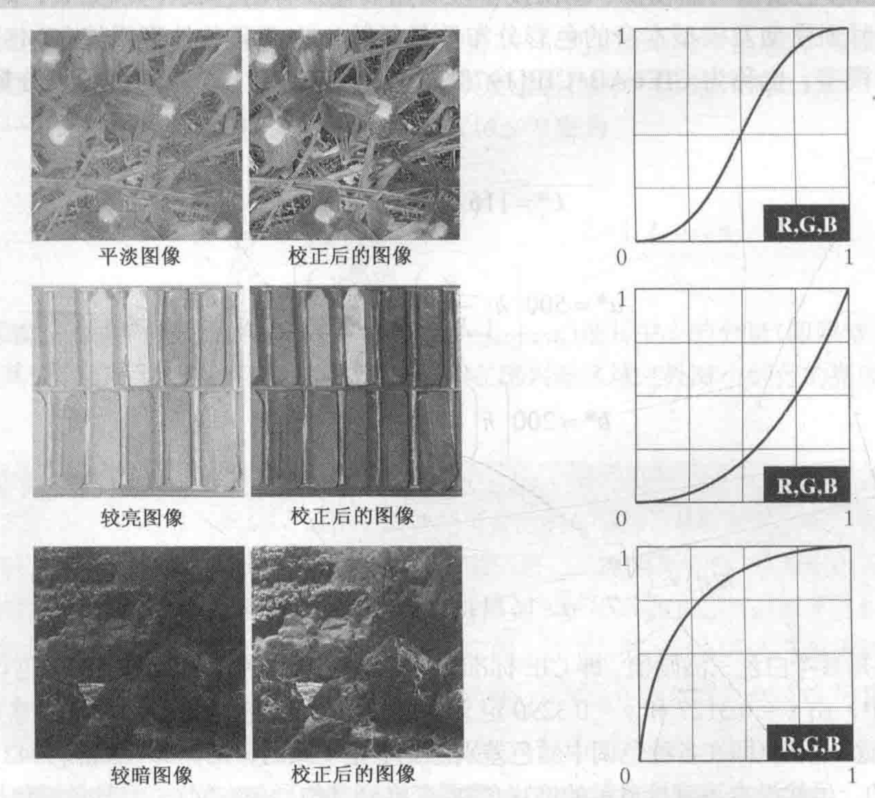


图 6.35　对平淡的、较亮的(高主调)、较暗的(低主调)彩色图像的色调校正。等量地调整红、绿、蓝分量并不总能显著地改变图像色调

例 6.10　彩色平衡。

正确地确立一幅图像的色调特性后，就可着手进行彩色平衡工作。虽然彩色平衡可以通过客观地(使用光谱仪)分析一幅图像中的已知颜色来确定，但当存在白色区域时，即这里 RGB 或 CMY(K)分量相等时，可以进行准确的视觉评估。如图 6.36 所示的那样，对于视觉颜色评估，皮肤色调是优秀的对象，因为人对固有的肤色高度敏感。鲜明的彩色，如亮红色的物体，在进行视觉彩色评估时并没有什么价值。

存在彩色不平衡时，校正它的方法有很多。在调整一幅图像的彩色分量时，要意识到每个操作都会影响到图像的全部彩色平衡。也就是说，对一种颜色的感知会受到周围颜色的影响。不过，图 6.32 的彩色环可用于预测一个彩色分量如何影响其他彩色分量。例如，基于该彩色环，任何颜色的比例都可通过减小图像中相对色(或补色)数量来增大。类似地，提高两种直接邻接颜色的比例，或减小与该补色邻接的两种颜色的

456 ~ 458

百分比，也可增大颜色的对比。例如，假定一幅 RGB 图像中有过多的深红色，则可通过(1)移去红色和蓝色或(2)增加绿色的方法减少它。

图 6.36 显示了用于校正简单 CMYK 输出不平衡的变换。注意，描述的变换是校正这些图像所要求的函数。这些函数的反函数用来产生相关的彩色不平衡。这些图像类似于暗室环境下的彩色印相，是识别彩色打印问题的有用参考工具。例如，过多的深红色(左下图像)或太少的青色(第二行最右侧的图像)会导致太多的红色。



原图像/校正后的图像

黑色较重　黑色较弱　青色较重　青色较弱

深红色较重　深红色较弱　黄色较重　黄色较弱

图 6.36　对 CMYK 图像的彩色平衡校正

6.5.5　直方图处理

与前一章中介绍的交互式增强方法不同，3.3 节的灰度直方图处理变换可自动地应用于彩色图像。回忆可知，直方图均衡会自动地确定一种变换，该变换试图产生具有均匀灰度值的直方图的图像。在单色图像情况下，业已证明(见图 3.20)在处理低、高和中主调图像时是合理而且成功的。然而，由于彩色图像是由多个分量组成的，所以必须考虑适应多个分量和/或直方图的灰度级技术。如预料的那样，单独对彩色图像的分量进行直方图均衡通常是不明智的，这将产生不正确的彩色。一种更合

458~460

乎逻辑的方法是，均匀地展开这种彩色灰度，而保持彩色本身(即色调)不变。下面的例子表明HSI彩色空间是适合这类方法的理想空间。

例6.11 HSI彩色空间中的直方图均衡。

图6.37(a)显示了含有调味瓶和摇杯的调味台的一幅彩色图像，其亮度分量(归一化后)的值域为[0, 1]。正如所看到的处理前的[见图6.37(b)]亮度分量直方图那样，图像包含有大量的暗彩色，使得中间灰度减少到0.36。不改变色调和饱和度，亮度分量直方图均衡的结果图像示于图6.37(c)中。注意，整个图像都被有效地加亮，并且一些调味瓶和放调味瓶的木桌的纹理现在都清晰可见。图6.37(b)显示了新图像的亮度直方图，以及用于均衡亮度分量的变换[见式(3.3-8)]。

虽然亮度均衡处理并不改变图像的色调和饱和度的值，但它的确影响了图像的整体颜色感观。特别要注意在不振动时，瓶中的油和醋。图6.37(d)显示了采用增加图像饱和度分量部分校正该问题的结果。然后使用图6.37(b)中的变换进行了直方图均衡。在处理HSI空间中的强度分量时，这种类型的调整很常见，因为灰度改变通常会影响图像中颜色的相对表观。

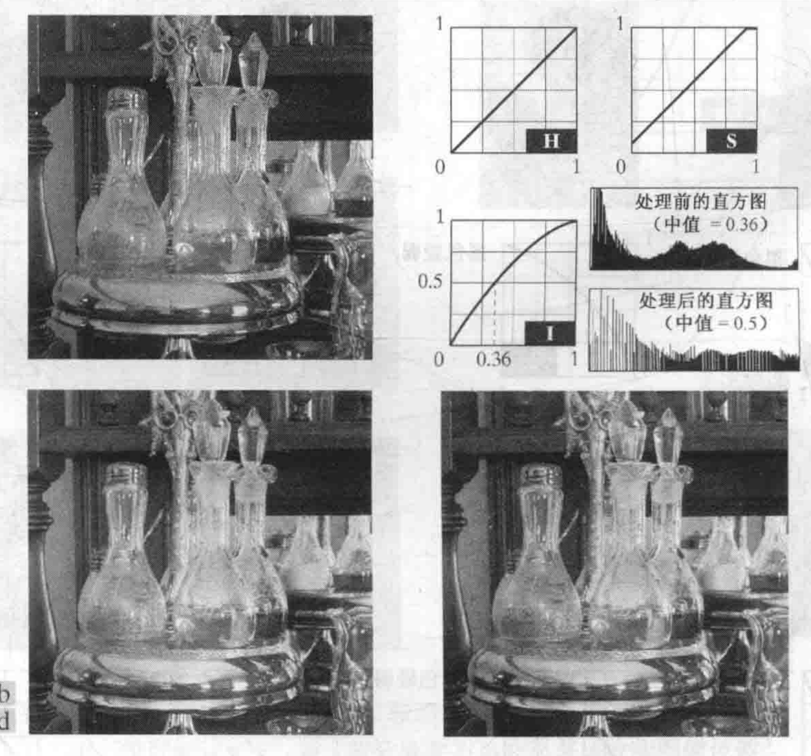


图6.37 HSI彩色空间中的直方图均衡(接着是饱和度调整)

6.6 平滑和锐化

在不涉及相邻像素的情况下(如前节那样)，对彩色图像的每个像素变换后的下一步是，以周围像素特性为基础改进该像素的值。本节通过彩色图像的平滑和锐化处理，说明这类邻域处理的基础。

6.6.1 彩色图像平滑

460~461

参考图6.29(a)和3.4节及3.5节的讨论，灰度级图像平滑可视为一种空间滤波操作，在这种操作中，滤波模板的系数具有相同的值。当模板滑过将被平滑的图像时，每个像素被由该模板定

义的邻域中的像素的平均值代替。如在图6.29(b)中看到的那样，这一概念很容易推广到全彩色图像处理。主要差别是，我们必须处理由式(6.4-2)给出的分量向量来替代灰度标量值。

在一幅 RGB 彩色图像中，令 S_{xy} 表示中心位于(x, y)的邻域定义的一组坐标。在该邻域中 RGB 分量的向量平均值为

$$\overline{c}(x,y)=\frac{1}{K}\sum_{(s,t)\in S_{xy}}c(s,t) \tag{6.6-1}$$

它满足式(6.4-2)和向量相加的性质，因此有

$$\overline{c}(x,y)=\begin{bmatrix}\dfrac{1}{K}\sum\limits_{(s,t)\in S_{xy}}R(s,t)\\ \dfrac{1}{K}\sum\limits_{(s,t)\in S_{xy}}G(s,t)\\ \dfrac{1}{K}\sum\limits_{(s,t)\in S_{xy}}B(s,t)\end{bmatrix} \tag{6.6-2}$$

我们将该向量的分量视为几幅标量图像，这些标量图像可通过传统的灰度级邻域处理，单独地平滑原 RGB 图像的每个平面来得到。这样，我们就可得出结论：邻域平均平滑可以在每个彩色平面的基础上执行。其结果与使用 RGB 彩色向量执行平均是相同的。

> 关于向量和矩阵的简要回顾，请查阅本书的网站。

例 6.12　用邻域平均法平滑彩色图像。

考虑图6.38(a)中的彩色图像，该图像的红、绿、蓝分量图像显示在图 6.38(b)到图 6.38(d)中。图 6.39(a)到图6.39(c)显示了该图像的 HSI 分量。基于前一段的讨论，我们用大小为 5×5 的空间平均模板独立地平滑图6.38 中的 RGB 图像的每个分量图像。然后，混合这些独立平滑后的图像以形成平滑后的全彩色结果，结果示于图 6.40(a)中。注意，正如在3.5 节中给出的几个例子那样，这幅图像显示了我们所希望的执行空间平滑操作的结果。

在6.2 节中，我们曾说明 HSI 彩色模型的一个重要优点是，解除了强度和彩色信息的联系，这就使得它适合于许多灰度处理技术，并且可能仅对图 6.39 中 HSI 描述的亮度分量平滑更有效。为了说明这一方法的优点和/或重要性，下面我们仅平滑亮度分量(保持色调和饱和度分量不变)，并为便于显示把处理结果转换为 RGB 图像。平滑后的彩色图像示于图6.40(b)中。注意，该图像类似于图6.40(a)，但正如从图 6.40(c)所示的差值图像中看到的那样，两幅平滑过的图像是不同的。这是因为在图6.40(a)中，每个像素的颜色是邻域中的像素的平均颜色。另一方面，通过仅平滑图6.40(b)中的亮度分量图像，每个像素的色调和饱和度不受影响，因此像素的颜色无变化。由这一观察可知，两种平滑方法间的差别会随着滤波器尺寸的增加而变得更为显著。






a b c d

图 6.38　(a)RGB 图像；(b)红分量图像；(c)绿分量图像；(d)蓝分量图像

461
~
463

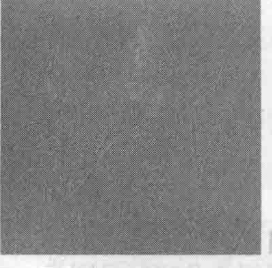




a b c

图 6.39 图 6.38(a)中的 RGB 彩色图像的 HSI 分量：(a)色调；(b)饱和度；(c)亮度





a b c

图 6.40 用 5×5 的平均模板平滑图像：(a)处理每一个 RGB 分量图像的结果；(b)处理 HSI 图像的亮度分量并转换为 RGB 图像后的结果；(c)两种结果的差别

6.6.2 彩色图像锐化

本节介绍采用拉普拉斯变换方法(见3.6.2节)的图像(尖)锐化处理。从向量分析可知，一个向量的拉普拉斯变换定义为一个向量，其分量等于输入向量的各个标量分量的拉普拉斯变换。在 RGB 彩色系统中，式(6.4-2)中向量 $\boldsymbol{c}$ 的拉普拉斯变换为

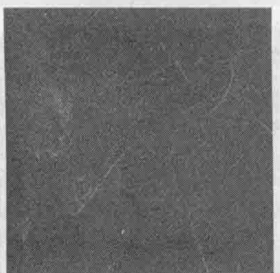
$$\nabla^2\left[\boldsymbol{c}(x,y)\right]=\begin{bmatrix}\nabla^2 R(x,y)\\ \nabla^2 G(x,y)\\ \nabla^2 B(x,y)\end{bmatrix} \tag{6.6-3}$$

如前节所述，它告诉我们可通过分别计算每幅分量图像的拉普拉斯变换来计算全彩色图像的拉普拉斯变换。

例 6.13 使用拉普拉斯变换的锐化。

图6.41(a)是使用式(3.6-7)和图3.37(c)中的模板计算图6.38 中的 RGB 分量图像的拉普拉斯变换得到的。这些结果合并在一起就产生了锐化后的全彩色图像。图6.41(b)显示了基于图 6.39 中 HSI 分量的类似的锐化图像。这一结果是通过合并亮度分量的拉普拉斯变换和未变化的色调与饱和度分量生成的。RGB 和 HSI 锐化图像间的差别示于图 6.41(c)。两幅图像之间存在这种差别的原因见例 6.12。





a b c

463 ~ 465

图 6.41 使用拉普拉斯变换锐化图像：(a)处理每个 RGB 通道的结果；(b)处理 HSI 亮度分量并转换为 RGB 图像后的结果；(c)两种结果的差别

6.7 基于彩色的图像分割

分割是把一幅图像分成多个区域的处理。虽然分割是第 10 章讨论的主题，但出于连续性的原因，这里简单地介绍一下彩色分割。在下面的讨论中，读者不会有理解方面的困难。

6.7.1 HSI 彩色空间的分割

如果我们希望以彩色为基础来分割一幅图像，并想在各个平面上执行这一处理，那么我们会很自然地首先想到 HSI 空间，因为在色调图像中描绘彩色很方便。通常，为了在色调图像中分离出感兴趣的孤立区域，我们会把饱和度用做一幅模板图像。在彩色图像分割中不常使用亮度图像，因为它不携带彩色信息。下面是如何在 HSI 彩色空间中执行分割的典型例子。

例 6.14 HSI 空间中的分割。

假定我们的兴趣是分割图 6.42(a) 中左下角图像中的微红色区。虽然它是用伪彩色方法产生的，但可以像全彩色图像那样处理(分割)而不失一般性。图 6.42(b) 到图 6.42(d) 是它的 HSI 分量图像。注意，比较图 6.42(a) 和图6.42(b) 发现，我们感兴趣的区域具有相对高的色调值，这表明这些颜色位于红色的蓝-深红侧(见图 6.13)。图 6.42(e) 显示了一个用阈值处理饱和度图像产生的二值模板，该阈值等于图像的最大饱和度的 10%。大于该阈值的任何像素值都被置为 1(白色)，所有其他值被置为 0(黑色)。

图6.42(f) 是色调图像与该模板的乘积，图6.42(g) 是乘积图像的直方图(注意，灰度级在范围[0, 1]内)。在直方图中可以看到，高值(感兴趣的值)被分组到灰度的最高端，接近 1.0。用阈值 0.9 对乘积图像进行阈值处理后得到的二值图像示于图6.42(h)。在该图像中，白色点的空间位置视为原图像的中点，它具有感兴趣的微红色调。这与完美的分割相差甚远，因为在原图像中还应该有一些具有微红色调的点，但它们未被该分割方法识别出来。然而，实验表明，在识别原图像的微红色分量时，图6.42(h) 所示的白色区域是这种方法所能做到的最好结果。下一节讨论的分割方法可产生相当好的结果。

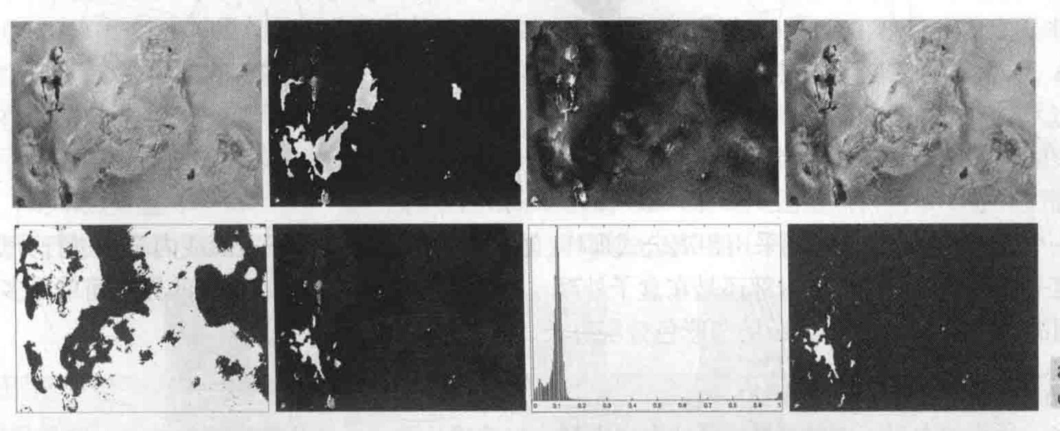


图 6.42　HSI 空间中的图像分割：(a) 原图像；(b) 色调图像；(c) 饱和度图像；(d) 亮度图像；(e) 二值饱和度模板(黑 = 0)；(f) 图 (b) 和图 (e) 相乘后的结果；(g) 图 (f) 的直方图；(h) 图 (a) 中红色分量的分割

6.7.2 RGB 向量空间中的分割

正如本节多次提到的那样，虽然在 HSI 空间的工作更直观，但分割却是用 RGB 彩色向量能得到更好结果的领域。这种方法很直接。假如我们的目的是在一幅 RGB 图像中分割某个指

465 ~ 467

定颜色区域的物体。给定一个感兴趣的有代表性彩色的彩色样点集,可得到我们希望分割的颜色的"平均"估计。令这个平均彩色用 RGB 向量 $\boldsymbol{a}$ 来表示。分割的目的是将给定图像中的每个 RGB 像素分类,即在指定的区域内是否有一种颜色。要执行这一比较,必须有一个相似性度量。最简单的度量之一是欧氏距离。令 $\boldsymbol{z}$ 表示 RGB 空间中的任意一点。如果它们之间的距离小于特定的阈值 D_0,则称 $\boldsymbol{z}$ 与 $\boldsymbol{a}$ 是相似的。$\boldsymbol{z}$ 和 $\boldsymbol{a}$ 间的欧氏距离由下式给出:

$$D(\boldsymbol{z},\boldsymbol{a})=\|\boldsymbol{z}-\boldsymbol{a}\|=\left[(\boldsymbol{z}-\boldsymbol{a})^{\mathrm{T}}(\boldsymbol{z}-\boldsymbol{a})\right]^{\frac{1}{2}}=\left[(z_R-a_R)^2+(z_G-a_G)^2+(z_B-a_B)^2\right]^{\frac{1}{2}} \tag{6.7-1}$$

其中下标 R, G, B 表示向量 $\boldsymbol{a}$ 和 $\boldsymbol{z}$ 的 RGB 分量。满足 $D(\boldsymbol{z},\boldsymbol{a})\leqslant D_0$ 的点的轨迹是半径为 D_0 的实心球体,如图 6.43(a)所示。包含在球体内部和表面上的点满足指定的颜色准则;球体之外的点则不满足指定的颜色准则。在图像中对这两组点编码,比如说黑或白,就产生了一幅二值分割图像。

式(6.7-1)的一种有用推广是形如下式的距离度量:

$$D(\boldsymbol{z},\boldsymbol{a})=\left[(\boldsymbol{z}-\boldsymbol{a})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{z}-\boldsymbol{a})\right]^{\frac{1}{2}} \tag{6.7-2}$$

其中 $\boldsymbol{C}$ 表示我们希望分割的有代表性颜色的样本的协方差矩阵[①]。满足 $D(\boldsymbol{z},\boldsymbol{a})\leqslant D_0$ 的点的轨迹描述了一个实心的三维椭球体[见图 6.43(b)],该椭球体的最大特点是主轴面向最大数据扩展方向。当 $\boldsymbol{C}=\boldsymbol{I}$ 时,则为 3×3 的单位矩阵,式(6.7-2)简化为式(6.7-1)。这时的分割与前一段中描述的分割相同。

因为距离是正的和单调的,所以可用距离的平方运算来代替,从而避免开方运算。然而,即使不计算平方根,对于实际大小的图像来说,实现式(6.7-1)或式(6.7-2)的计算代价也很高。一种折中方案是使用一个边界盒,如图6.43(c)所示。在该方法中,盒的中心在 $\boldsymbol{a}$ 处,沿每个颜色轴选择的维数与沿每个轴的样本的标准差成比例。标准差的计算只使用一次样本颜色数据。

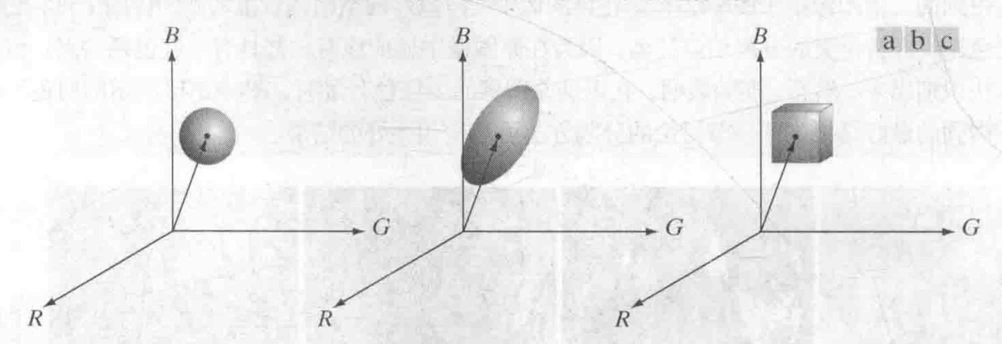


图 6.43 为 RGB 向量分割圈定数据区域的 3 种方法

给定一个任意的彩色点,如同采用距离公式那样,通过确定它是否在盒子表面或内部来进行分割。然而,判定一个彩色点是在盒子内部还是在盒子外部,在计算上要比求球体或椭球的界限简单得多。注意,前面的讨论是 6.5.3 节中介绍的与彩色分层有关方法的推广。

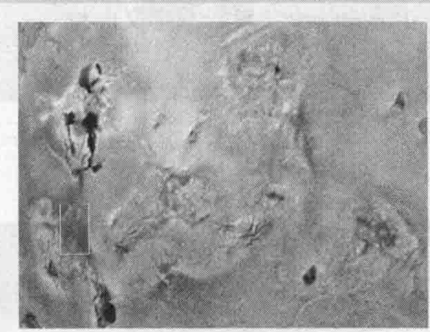
例 6.15 RGB 空间中的彩色图像分割。

图 6.44(a)所示的方形区域包含微红色样本,它是我们希望从彩色图像分割出来的样本。这与我们在例6.14中使用色调的问题相同,但这里我们使用RGB彩色向量来处理该问题。接下来的方法是使用图6.44(a)所示矩形内包含的彩色点来计算平均向量 $\boldsymbol{a}$,然后计算这些样本的红、绿、蓝值的标准差。盒子的中心位于 $\boldsymbol{a}$ 处,其沿每个 RGB 轴的维数按沿相应轴的数据的标准差的 1.25 倍选取。例如,令 σ_R 表示样本点的红分量的标准差,则盒子沿 R 轴的维数从 $(a_R-1.25\sigma_R)$ 扩展到 $(a_R+1.25\sigma_R)$,其中 a_R 表示平均向量 $\boldsymbol{a}$ 的红分量。

467 ~ 468

① 一组向量样本的协方差矩阵的计算,见 11.4 节中的讨论。

对整个彩色图像中的每一点的编码结果如下：如果该点位于盒子的表面或内部，则编码为白色；否则编码为黑色。编码后的结果示于图6.44(b)。注意分割后的区域是如何由矩形封闭的彩色样本推广的。事实上，比较图6.44(b)和图6.42(h)，可看到在RGB向量空间分割会产生准确得多的结果。在一定意义上，在原彩色图像中它们更接近于我们定义的“微红色”。





a b

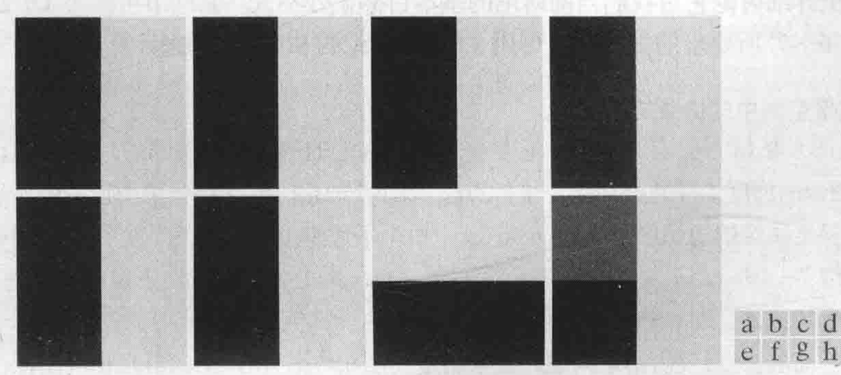
图6.44　RGB空间中的分割：(a)用一个封闭的矩形表示的感兴趣的彩色原像；(b)RGB向量空间中的分割结果。请与图6.42(h)比较

6.7.3　彩色边缘检测

如第10章讨论的那样，边缘检测对图像分割来说是一个重要的工具。在本节中，我们的兴趣在于以单一图像为基础计算边缘和直接在彩色向量空间中计算边缘的问题。基于边缘的分割的详细内容将在10.2节中给出。

梯度算子边缘检测已在3.6.4节有关图像锐化的内容中介绍过。遗憾的是，3.6.4节中讨论的梯度对向量没有定义。这样，我们会立刻想到，计算单幅图像的梯度，然后用得到的结果形成一幅彩色图像，将会导致错误的结果。一个简单的例子可帮助我们说明其原因。

考虑图6.45(d)和图6.45(h)中的两幅$M \times M$彩色图像(M为奇数)，它们分别由图6.45(a)到图6.45(c)和图6.45(e)到图645(g)中的3个分量图像组成。例如，如果计算每个分量图像的梯度[见式(3.6-11)]，并将结果相加形成两幅相应的RGB梯度图像，则点$[(M+1)/2, (M+1)/2]$处的梯度值在两种情况下都将相同。直观地看，我们希望图6.45(d)中图像那一点的梯度更强，因为R, G, B图像的边缘在该图像中的方向相同，而相对于图6.45(h)的图像，只有两个边缘具有相同的方向。从这个简单的例子可以看出，处理3个单独的平面所形成的合成梯度图像可导致错误的结果。如果问题只是检测边缘，则单独分量方法通常可产生能接受的结果。然而，如果准确度是重点，则很明显我们需要一个适用于向量的梯度的新定义。下面讨论Di Zenzo[1986]为此提出的一种方法。



a b c d
e f g h

图6.45　(a)~(c)R, G, B分量图像；(d)产生的RGB彩色图像；(e)~(g)R, G, B分量图像；(h)产生的RGB彩色图像

468~470

现有的问题是定义式(6.4-2)中向量 $\boldsymbol{c}$ 在任意点(x, y)处的梯度(幅度和方向)。如刚提及的那样，3.6.4节中我们研究过的梯度适用于标量函数 $f(x,y)$，但不能用于向量函数。下面是许多方法中的一种，在这种方法中，我们可以针对向量函数扩展梯度的概念。回忆可知，对于标量函数$f(x,y)$，梯度是坐标(x, y)处指向f的最大变化率方向的向量。

令 $\boldsymbol{r}, \boldsymbol{g}$ 和 $\boldsymbol{b}$ 是沿 RGB 彩色空间(见图 6.7)的 R, G, B 轴的单位向量，并定义向量

$$\boldsymbol{u} = \frac{\partial R}{\partial x}\boldsymbol{r} + \frac{\partial G}{\partial x}\boldsymbol{g} + \frac{\partial B}{\partial x}\boldsymbol{b} \tag{6.7-3}$$

和

$$\boldsymbol{v} = \frac{\partial R}{\partial y}\boldsymbol{r} + \frac{\partial G}{\partial y}\boldsymbol{g} + \frac{\partial B}{\partial y}\boldsymbol{b} \tag{6.7-4}$$

令 g_{xx}，g_{yy} 和 g_{xy} 表示这些向量的点积，如下所示：

$$g_{xx} = \boldsymbol{u}\cdot\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial G}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2 \tag{6.7-5}$$

$$g_{yy} = \boldsymbol{v}\cdot\boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial G}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2 \tag{6.7-6}$$

和

$$g_{xy} = \boldsymbol{u}\cdot\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial G}{\partial x}\frac{\partial G}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y} \tag{6.7-7}$$

记住，R, G 和 B 以及由此而来的 g 项是 x 和 y 的函数。使用这种表示法，可以证明(Di Zenzo[1986])，$\boldsymbol{c}(x, y)$的最大变化率方向可以由角度

$$\theta(x, y) = \frac{1}{2}\arctan\left[\frac{2g_{xy}}{g_{xx} - g_{yy}}\right] \tag{6.7-8}$$

给出，且在角度$\theta(x, y)$方向上点(x, y)处的变化率的值由下式给出：

$$F_\theta(x, y) = \left\{\frac{1}{2}\left[(g_{xx} + g_{yy}) + (g_{xx} - g_{yy})\cos 2\theta(x, y) + 2g_{xy}\sin 2\theta(x, y)\right]\right\}^{\frac{1}{2}} \tag{6.7-9}$$

因为$\tan(\alpha) = \tan(\alpha \pm \pi)$，如果$\theta_0$是式(6.7-8)的一个解，则$\theta_0 \pm \pi/2$也是该式的一个解。此外，由于$F_\theta = F_{\theta+\pi}$，所以$F$仅需对$\theta$值在半开区间$[0, \pi)$内计算。式(6.7-8)提供两个相隔 90°的值这一事实，意味着该式在每个点(x, y)处涉及两个正交方向。沿着这些方向之一，F最大，沿其他方向F最小。这些结果的推导相当冗长，并且在这里详细讨论它对我们当前讨论的基本目标益处不大。其细节可参考 Di Zenzo[1986]。实现式(6.7-5)到式(6.7-7)所要求的偏导数，可用 3.6.4 节讨论的 Sobel 算子来计算。

例 6.16　向量空间中的边缘检测。

图6.46(b)是图 6.46(a)中图像的梯度，它是使用刚刚讨论的向量方法得到的。图 6.46(c)显示了由计算每幅 RGB 分量图像的梯度，并通过在每一坐标点(x, y)处叠加相应的 3 幅分量图像的值所形成的一幅合成梯度图像。向量梯度图像的边缘细节比图 6.46(c)中单个平面梯度图像的细节更完整。例如，可观察研究对象右眼周围的细节。图6.46(d)中的图像显示了在每一点(x, y)处的两幅梯度图像间的差别。注意，两种方法都得到了可接受的结果。图 6.46(b)中额外细节所增加的计算负担是否值得(与 Sober 算子的执行相反，它用于产生单一平面的梯度)仅由给定问题的要求决定。图6.47 显示了 3 幅分量梯度图像，在相加和标定后可用来得到图 6.46(c)。

470 ~ 472






a b c d

图 6.46　(a) RGB 图像；(b) 在 RGB 彩色空间中计算的梯度；(c) 在每幅图像基础上计算梯度，然后相加；(d) 图 (b) 和图 (c) 的差





a b c

图 6.47　图 6.46 中的彩色图像的分量梯度图像：(a) 红分量；(b) 绿分量；(c) 蓝分量。这 3 幅图像相加并标定后产生了图 6.46 (c) 中的图像

6.8　彩色图像中的噪声

5.2 节中讨论的噪声模型可用于彩色图像。通常，彩色图像的噪声内容在每个彩色通道中具有相同的特性，但噪声对不同彩色通道所造成的影响不同。一种可能是个别通道的电子学故障。然而，不同的噪声水平像是由每个彩色通道的相对照射强度的差异造成的。例如，在 CCD 摄像机中，红色滤镜的使用将减小用于红色传感器的照射强度。CCD 传感器在低照明水平下就是噪声源，因此，这种情况下得到的 RGB 图像的红色分量图像与其他两幅分量图像相比往往是噪声源。

例 6.17　将 RGB 带噪图像转换为 HSI 的效果说明。

在这个例子中，我们将简单地看一下彩色图像中的噪声，以及在从一个彩色模型转换为另一个模型时，噪声是如何转移的。图 6.48 (a) 到图 6.48 (c) 显示了被高斯噪声污染的一幅 RGB 图像的 3 个彩色平面，图 6.48 (d) 是合成的 RGB 图像。注意，与在单色图像中相比，细粒噪声在彩色图像中不太引人注意。图 6.49 (a) 到图 6.49 (c) 显示了把图 6.48 (d) 中的 RGB 图像转换为 HSI 的结果。将这些结果与原图像（见图 6.39）的 HSI 分量进行比较，会发现噪声图像的色调与饱和度分量明显降低了。这分别是由式(6.2-2)和式(6.2-3)求余弦与取最小值操作的非线性造成的。另一方面，图6.49 (c) 中的强度分量比 3 个带噪声的 RGB 分量图像中的任何一个都要平滑一些。如式 (6.2-4) 指出的那样，这是由亮度图像是 RGB 图像的平均这一事实造成的（回忆 2.6.3 节中关于对图像进行平均操作降低随机噪声的讨论）。

在仅有一个 RGB 通道受噪声影响的情况下，到 HSI 的转换才将噪声扩散到所有 HSI 分量图像。图 6.50 (a) 给出了一个例子。图6.50 (a) 显示了一幅 RGB 图像，其绿色图像被椒盐噪声污染，其中无论是盐粒噪声还是胡椒噪声的概率都为 0.05。图6.50 (b) 到图6.50 (d) 中的 HSI 分量图像清楚地显示了噪声是如何从绿色 RGB 通道散布到所有 HSI 图像上的。当然，这是我们不希望的，因为 HSI 分量的计算会用到 RGB 的所有分量，正如 6.2.3 节所示的那样。

472 ~ 475






a b c d

图 6.48 (a)~(c)由均值为 0、方差为 800 的高斯噪声污染的红、绿和蓝分量图像；(d)最终的 RGB 图像［与图 6.46(a)比较］





a b c

图 6.49 图 6.48(d)中带噪声的彩色图像的 HSI 分量：(a)色调图像；(b)饱和度图像；(c)亮度图像






a b c d

图 6.50 (a)绿平面被椒盐噪声污染的 RGB 图像；(b)HSI 图像的色调分量；(c)饱和度分量；(d)亮度分量

迄今为止我们讨论过的处理都是正确的，因此全彩色图像滤波可以在每幅图像的基础上执行，或依靠该过程直接在彩色向量空间中执行。例如，用均值滤波器减少噪声是 6.6.1 节讨论过的处理方法，我们知道，在向量空间中对分量图像单独进行类似的处理时，可以给出相同的结果。然而，其他滤波器却不能按这样的方式明确表达。这样的例子包括 5.3.2 节中讨论过的各类排序统计滤波。例如，要在彩色向量空间实现中值滤波，就要找到一种使得中值有意义的向量排序方案。处理标量的过程很简单，而处理向量的过程相当复杂。向量排序超出了此处的讨论范围，但 Plataniotis and Venetsanopoulos[2000]是关于向量排序和某些基于排序概念的滤波的较好参考书。

6.9 彩色图像压缩

因为描述彩色所要求的比特数比描述灰度级所要求的比特数大 3~4 倍，所以数据压缩在存储和传输彩色图像中起着核心的作用。对于前几节中描述的 RGB, CMY(K)和 HSI 图像，任何压缩对象的数据是每个彩色像素的分量(即 RGB 图像中各像素的红、绿、蓝分量)；它们是传达彩色信息的方式。压缩是减小或消除冗余和/或不相干数据的处理。虽然压缩是第 8 章的主题，但我们在下面的例子中用一幅彩色图像来简单地说明这一概念。

475~476

例 6.18　彩色图像压缩实例。

图6.51(a)显示了蝴蝶花的一幅 24 比特 RGB 全彩色图像，图像中表示红、绿、蓝分量时都使用了 8 比特。图 6.51(b)是由图 6.51(a)所示图像的一种压缩方案重建的图像，事实上，该图像是压缩并解压缩后的一种近似。虽然压缩后的图像不能直接显示——在输入到彩色监视器之前必须解压缩，但对于原图像中每 230 比特的数据，压缩后的图像仅包含 1 个数据比特（只需要 1 个存储比特）。假定压缩后的图像在 Internet 上能于 1 分钟内传输完成，那么原图像的传输至少需要 4 小时。当然，传输后的数据必须解压缩后才能观看，但解压缩在几秒之内就可以完成。用于生成图 6.51(b)的 JPEG 2000 压缩算法是近来引入的标准，该标准将在 8.2.10 节中详细叙述。注意，重建的近似图像稍微有点模糊。这是许多有损压缩技术的特点；我们可以通过改变压缩程度将模糊减小或消除。




a b

图 6.51　彩色图像压缩：(a)原 RGB 图像；(b)图(a)压缩并解压缩后的图像

小结

本章简单地介绍了彩色图像处理，所涵盖的主题为图像处理技术提供了坚实的背景知识。彩色基本原理和彩色模型是彩色图像处理领域的基本内容，这些内容在技术范围和应用领域方面非常宽泛。我们特别关注了彩色模型，它们不仅在数字图像处理中非常有用，而且还为图像处理领域的深入研究提供了必要的工具。单幅图像的伪彩色和全彩色处理的讨论，为第 3 章到第 5 章中的某些技术提供了纽带。

在探讨彩色向量空间时，首先介绍了以前研究的方法，突出了灰度和全彩色处理间的重要区别。就技术而论，直接进行彩色向量处理的领域很多，且包括许多方法，如中值滤波和其他排序滤波、自适应和形态学滤波、图像复原、图像压缩等。这些处理不同于在彩色图像的各幅分量图像上执行的彩色处理。下节中的参考文献提供了该领域的深入研究成果。

在彩色图像的噪声处理方面，我们遵循彩色图像从一种工作空间变换至另一种工作空间的事实，指出了问题的向量性质，以及降低噪声的方法。在某些情况下，噪声的滤除可以在每幅图像的基础上进行；但另一方面，如中值滤波，就要求特殊处理，以便反映彩色像素是向量这一事实，就像前几段中提到的那样。

虽然分割是第 10 章的主题，图像数据压缩是第 8 章的主题，但在本章中介绍它们有利于我们保持知识的连续性。后续的讨论将表明，这些章节中介绍的许多技术同样适用于本章的讨论。

参考文献

色彩学的全面参考见 Malacara[2001]。关于彩色生理学，请参阅 Gegenfurtner and Sharpe[1999]。这两个参考文献与 Walsh[1958]和 Kiver[1965]一起，为 6.1 节的讨论提供了丰富的补充材料。彩色模

476 ~ 478

型(见 6.2 节)的深入读物见 Fortner and Meyer[1997]、Poynton[1996]和 Fairchild [1998]。6.2.3 节中 HSI 模型的详细推导，见 Smith[1978]，也可查阅本书的网站。伪彩色这一主题(见 6.3 节)与数据可视化这一通用领域联系紧密。在伪彩色的应用方面，Wolff and Yaeger[1993]是很好的基础参考文献，Thorell and Smith[1990]也很重要。关于彩色信号向量表示的讨论(6.4 节)，见 Plataniotis and Venetsanopoulos [2000]。

6.5 节的参考文献是 Benson[1985]、Robertson[1977]和 CIE[1978]，还可参阅 MacAdam[1942]。彩色图像滤波方面的内容(见 6.6 节)是以 6.4 节介绍的向量公式及第 3 章讨论的空间滤波为基础的。彩色图像分割(见 6.7 节)是过去 10 年间备受关注的主题。Liu and Yang[1994]和 Shafarenko et al.[1998]是这一领域的代表性文献。IEEE 图像处理学报的专刊[1997]也是重要的参考资料。彩色边缘检测的讨论(见 6.7.3 节)来自 Di Zenzo[1986]。Plataniotis and Venetsanopoulos[2000]在综述彩色图像分割的各种方法方面做了很好的工作。6.8 节的讨论以 5.2 节中介绍的噪声模型为基础。图像压缩(见 6.9 节)方面的文献列在第 8 章的末尾。本章讨论的许多技术的软件实现见 Gonzalez, Woods and Eddins[2004]。

> 标有星号的习题的详细解答，可在本书的网站上找到。该网站还包含有基于本章内容建议的项目。

习题

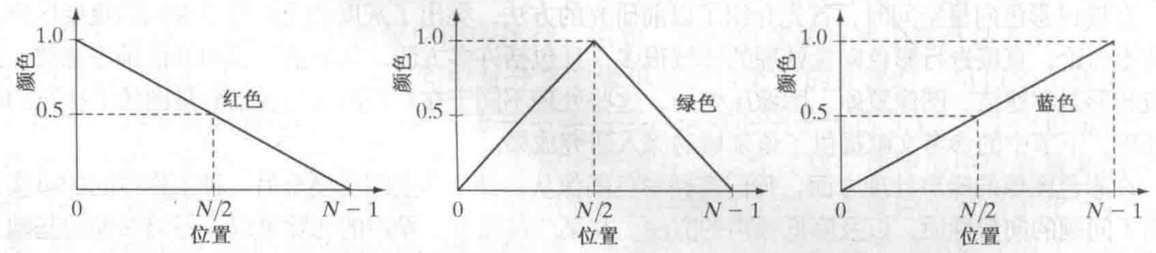
6.1 给出用于产生图 6.5 中标为“日光”的点的红光、绿光、蓝光的百分比。

★6.2 考虑任何两种有效颜色 c_1 和 c_2，其在图 6.5 所示色度图中的坐标分别为(x_1, y_1)和(x_2, y_2)。推导出计算构成一种给定颜色的 c_1 和 c_2 的相对百分比的通用表示式，已知给定的颜色位于连接这两种颜色的直线上。

6.3 考虑任何 4 种有效彩色 c_1, c_2, c_3 和 c_4，其在图 6.5 所示色度图中的坐标为(x_1, y_1), (x_2, y_2), (x_3, y_3)和(x_4, y_4)。推导出计算构成给定彩色的 c_1, c_2, c_3 和 c_4 的相对百分比的通用表达式，已知给定的颜色位于由 4 个项点 c_1, c_2, c_3 和 c_4 构成的方形内。

★6.4 在一条自动装配线中，为了简化检测，四类零件将编以彩色码。但只有一台黑白电视摄像机可用于获取数字图像。请提出一种使用该摄像机检测 4 种不同颜色的技术。

6.5 在一幅简单的 RGB 图像中，R, G 和 B 分量图像有下图所示的水平强度分布曲线。人们在该图像的中间一列能看到什么颜色？



★6.6 当它们出现在单色监视器上时，画出右侧图像的 RGB 分量。所有颜色都处在最大亮度与饱和度。在解答这一习题时，考虑作为图像一部分的中间灰度边界。

6.7 在彩色 RGB 系统中，每幅 RGB 图像均为 12 比特图像，共有多少种不同的灰度？

6.8 考虑图 6.8 所示的 RGB 彩色立方体并回答下列问题：

★(a) 描述在 R，G 和 B 原图像中灰度级是如何变化的。原图像由彩色立方体的前表面组成。

(b) 假设用 CMY 彩色代替 RGB 立方体中的每种颜色，这个新立方体显示在 RGB 监视器上。请用彩色名标注在屏幕上看到的新立方体的 8 个顶点。

478 ~ 479

(c) 请说明与饱和度有关的 RGB 彩色立方体边缘上的彩色。

6.9 (a) 当习题 6.6 中的图像出现在单色监视器上时，请画出该图像的 CMY 分量。

(b) 如果(a)中画出的 CMY 分量分别送入监视器的红、绿、蓝输入端，试描述结果图像。

★6.10 从式(6.5-5)中的 RGB 亮度映射函数推导出式(6.5-6)的 CMY 亮度映射函数。

6.11 考虑如图 6.10(a)所示的整个 216 种稳定色阵列。按(行，列)方式标出每一元素，左上角元素为(1, 1)，右下角元素为(12, 18)。请问在哪个元素会找到

(a) 最纯的红色？

(b) 最纯的黄色？

★6.12 当习题 6.6 中图像的 HSI 分量出现在单色监视器上时，画出这些分量。

6.13 提供一种方法，以生成一个与图 6.2 中称为“可见光谱”的放大部分相似的彩色波段。注意该波段左边从暗紫色开始，并向右边的纯红色延续。(提示：使用 HSI 彩色模型。)

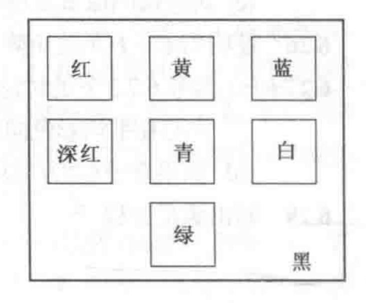
★6.14 提出一种生成图 6.13(c)中以图形方式显示的图像的彩色方案的方法，并以流程图的方式给出答案。假定亮度值固定且已给定。(提示：使用 HSI 彩色模型。)

6.15 考虑右下方由纯色方形组成的图像。为讨论你的答案，选择由 0~7 的 8 个灰度组成的灰度级，其中 0 表示黑，7 表示白。假设该图像转换到 HSI 彩色空间。如果使用的数字有意义，在回答如下问题时，对灰度使用一个明确的数字。否则，使用“与……相同”、“比……亮”或“比……暗”的关系就已足够。如果不能对正讨论的图像赋予一种特定的灰度级或这些关系之一，试给出原因。

(a) 画出色调图像的草图。

(b) 画出饱和度图像的草图。

(c) 画出亮度图像的草图。

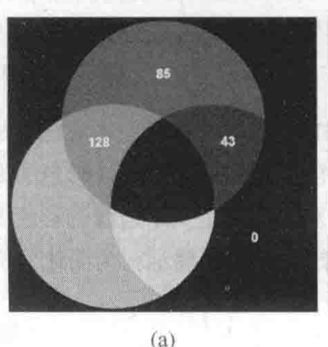


6.16 下列 8 比特图像(从左到右)是来自图 6.16 的 H, S, I 分量图像。数字指出了灰度值。回答下列问题，解释回答的依据。如果基于给定的信息不可能回答这一问题，则陈述为什么不能回答。

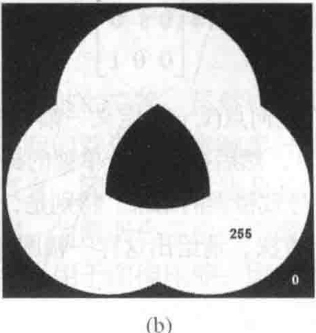
★(a) 给出色调图像中所有区域的灰度值。

(b) 给出饱和度图像中所有区域的灰度值。

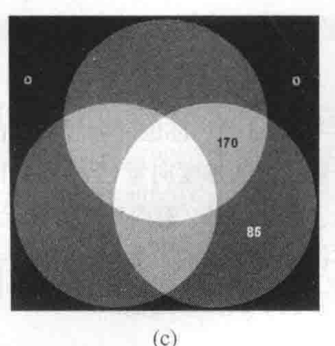
(c) 给出亮度图像中所有区域灰度值。



(a)



(b)



(c)

6.17 参考图 6.27 回答下列问题：

(a) 为什么图 6.27(f)中的图像主要呈现红色调？

(b) 提出一种自动编码程序，使图 6.27 中的水为亮蓝色。

(c) 提出一种自动编码程序，使主要的人工分量呈现亮红色。[提示：使用图 6.27(f)。]

★6.18 证明彩色图像的补色的饱和度分量，不能单独地由输入图像的饱和度分量计算出来。

6.19 使用 HSI 彩色模型说明图 6.33(b)中的近似补色的色调变换函数的形状。

480~481

★6.20 推导产生一幅彩色图像的补色的CMY变换。

6.21 画出用于校正RGB彩色空间中过度对比度的变换函数的一般形状。

★6.22 假定一个成像系统的监视器和打印机未完美校准。在该监视器上看起来平衡的一幅图像打印时出现了青色。描述可校正这种不平衡的通用变换。

6.23 假定

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.588 & 0.179 & 0.183 \\ 0.29 & 0.606 & 0.105 \\ 0 & 0.068 & 1.021 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

计算习题6.6中图像的$L^*a^*b^*$分量。该矩阵方程定义了由(美国)国家电视标准委员会(NTSC)彩色电视荧光粉产生的在D65标准照明下观看的彩色三色激励值(Benson[1985])。

★6.24 怎样才能实现与3.3.2节中匹配(规定)的灰度直方图等效的彩色直方图?

6.25 考虑下列500×500的RGB彩色图像，图像中方框是完全饱和的红色、绿色和蓝色，且每种颜色处在最大亮度[例如，对于红色方框是(1, 0, 0)]。由该图像生成了一幅HSI图像。

(a) 描述每幅HSI分量图像的外观。

(b) 用一个250×250的平均模板来平滑HSI图像的饱和度分量。描述结果的外观(可忽略滤波操作对图像边界的影响)。

(c) 对色调图像重复(b)。

绿色	红色
蓝色	绿色

6.26 证明当$\boldsymbol{C} = \boldsymbol{I}$(单位矩阵)时，式(6.7-2)简化为式(6.7-1)。

6.27★(a) 参考6.7.2节的讨论，(以流程图形式)给出一个过程，确定彩色向量(点)$\boldsymbol{z}$是否在一个边长为W、中心在平均彩色向量$\boldsymbol{a}$处的立方体内部。不允许计算距离。

(b) 如果盒子(立方体)与坐标轴是对齐的，那么可对各幅图像进行这一处理。说明你应该怎么做。

6.28 画出满足方程

$$D(\boldsymbol{z}, \boldsymbol{a}) = [(\boldsymbol{z} - \boldsymbol{a})^{\mathrm{T}} \boldsymbol{C}^{-1} (\boldsymbol{z} - \boldsymbol{a})]^{1/2} = D_0$$

的那些点在RGB空间中组成的表面，其中D_0是一个指定的非零常数。假定$\boldsymbol{a} = 0$且

$$\boldsymbol{C} = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

6.29 参考6.7.3节，你也许会想到，在任何点(x, y)处定义一幅RGB图像的梯度的逻辑方法，是计算每幅分量图像的梯度向量(见3.6.4节)，然后把这3个单独的梯度向量相加，形成该彩色图像的梯度向量。遗憾的是，有时这一方法会导致错误的结果。特别是，对于明确定义了边缘的彩色图像，具有零梯度是可能的。如果使用这一方法，请给出这样一幅图像的例子。(提示：为简化分析，可置一个彩色平面为恒定值。)

第7章 小波和多分辨率处理

All this time, the guard was looking at her, first through a telescope, then through a microscope, and then through an opera glass.

Lewis Carrol, Through the Looking Glass

引言

虽然自20世纪50年代末起，傅里叶变换就一直是基于变换的图像处理的基石，但近年来一种新的称为小波变换的变换使得压缩、传输和分析图像变得更为容易。与基函数为正弦函数的傅里叶变换不同，小波变换的基函数是小型波，称为小波，它具有变化的频率和有限的持续时间。这就使得它们为图像提供一张等效的乐谱，该乐谱不但显示了要演奏的音符(或频率)，而且显示了演奏这些音符的时间。另一方面，傅里叶变换只提供音符或频率信息，时间信息在变换过程中则丢失了。

1987年，人们首次证明小波是*多分辨率理论*这种信号处理和分析方法的基础(Mallat [1987])。多分辨率理论有效地统一了多个学科的技术，包括来自信号处理的子带编码、来自数字语音识别的正交镜像滤波及金字塔图像处理。如其名称所示，多分辨率理论涉及多个分辨率下的信号(或图像)表示与分析。这种方法的优势很明显，某种分辨率下无法检测的特性，在另一种分辨率下会很容易检测。尽管20世纪80年代末之前，图像处理界对多分辨率分析的研究兴趣有限，但现在关于该专题的论文和专著却多得数不胜数。

本章从多分辨率的角度来审视基于小波的变换。虽然这样的变换也可使用其他方法来介绍，但这种方法能简化它们的数学和物理解释。我们首先概述影响多分辨率理论形成的图像处理技术，以便在图像处理环境中阐述这一理论的基本概念，同时对该方法及其应用进行简要的历史回顾。本章的主要内容集中于离散小波变换的开发和利用。为阐明小变换的用途，本章提供了图像编码、噪声去除和边缘检测等例子。在下一章中，小波将主要用于图像压缩，因为小波在图像压缩中的应用已得到了人们的充分重视。

7.1 背景

我们观察图像时，看到的通常是由相似纹理和灰度级连成的区域，它们相结合就形成了物体。如果物体的尺寸较小或对比度较低，那么我们通常以较高的分辨率来研究它们；如果物体的尺寸较大或对比度较高，则粗略的观察就已足够。如果较小物体和较大物体（或对比度较低和对比度

483
~
484

较高的物体）同时存在，那么以不同分辨率来研究它们将更具优势。当然，这就是多分辨率处理的基本动机。

从数学角度来看，图像是具有局部变化统计特性灰度值的二维阵列，而这种统计特性是由类似边缘和对比同质区域突变特性的不同组合导致的。如图7.1所示（本章剩余部分将多次使用这一图像），在同一幅图像的不同部分，局部直方图可能变化很明显，这就使得为整幅图像建立统计模型非常困难，甚至不可能建立统计模型。

局部直方图是一个邻域中的像素的直方图(见3.3.3节)。

7.1.1 图像金字塔

以多个分辨率来表示图像的一种结构是图像金字塔，这种结构非常有效，且概念简单(Burt and Adelson[1983])。图像金字塔最初用于机器视觉和图像压缩，是一系列以金字塔形状排列的、分辨率逐步降低的图像集合。就像在图7.2(a)中看到的那样，金字塔的底部是待处理图像的高分辨率表示，而顶部则包含一个低分辨率近似。向金字塔上层移动时，尺寸和分辨率逐步降低。基础级 J 的大小为 $2^J \times 2^J$ 或 $N \times N$，其中 $J = \log_2 N$，顶点级0的大小为1×1(即单个像素)，通常第 j 级的大小为 $2^j \times 2^j$，其中 $0 \leqslant j \leqslant J$。虽然图7.2(a)所示的金字塔由从 $2^J \times 2^J$ 到 $2^0 \times 2^0$ 的 $J+1$ 个分辨率级别组成，但大部分图像金字塔会截短到 $P+1$ 级，其中 $1 \leqslant P \leqslant J$ 且 $j = J-P, \cdots, J-2, J-1, J$。也就是说，我们通常会将级别限制到 P 来降低原图像的分辨率近似；例如，一幅大小为512×512的图像的1×1(即单像素)近似没有什么价值。$P+1$ 级金字塔($P>0$)中的像素总数是

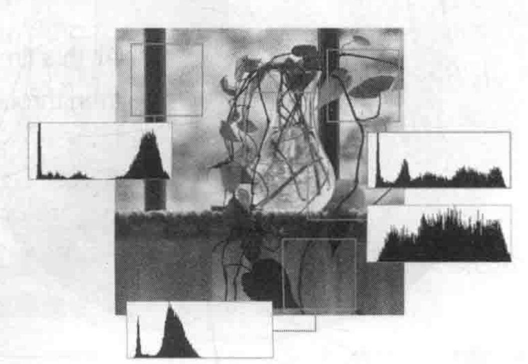


图7.1　一幅图像及其不同的局部直方图

$$N^2\left(1+\frac{1}{(4)^1}+\frac{1}{(4)^2}+\cdots+\frac{1}{(4)^P}\right) \leqslant \frac{4}{3}N^2$$

图7.2(b)显示了一个简单的系统，该系统构了两个紧密联系的图像金字塔。第 $j-1$ 级近似输出提供建立一个近似值金字塔所需的图像(如前一段中描述的那样)，第 j 级预测残差输出构建一个补充的预测残差金字塔。与近似金字塔不同，预测残差金字塔仅包含输入图像的一个分辨率降低的近似(在金字塔的顶部，即 $J-P$ 级)。所有其他级均包含预测残差，其中第 j 级预测残差(对于 $J-P+1 \leqslant j \leqslant J$)定义为第 j 级近似(方框图的输入)与基于第 $j-1$ 级近似(方框图中的近似输出)的第 j 级近似的估计之间的差。

一般而言，预测残差可定义为一幅图像与该幅图像的一个预测版本间的差。如8.2.9节将说明的那样，与二维灰度阵列相比，预测残差可更有效率地编码。

如图7.2(b)建议的那样，近似和预测残差金字塔都以一种迭代方式进行计算。在第一次迭代前，以金字塔形式描绘的图像放在这个近似金字塔的第 J 级中。然后，下面的3步过程执行 P 次，其中 $j = J, J-1, \cdots, J-P+1$：

步骤1　计算第 j 级输入图像分辨率降低的近似[图7.2(b)所示方框图左侧的输入]。这可通过滤波并对滤波后的结果以2为基进行下采样来完成。这两种操作将在下一段中描述。将得到的近似放在近似金字塔的第 $j-1$ 级。

步骤2　由步骤1产生的分辨率降低的近似，创建第 j 级输入图像的一个估计。方法是对产生的近似与第 j 级图像进行上采样和滤波(见下节)。得到的预测图像与第 j 级输入图像的维数相同。

484
~
486

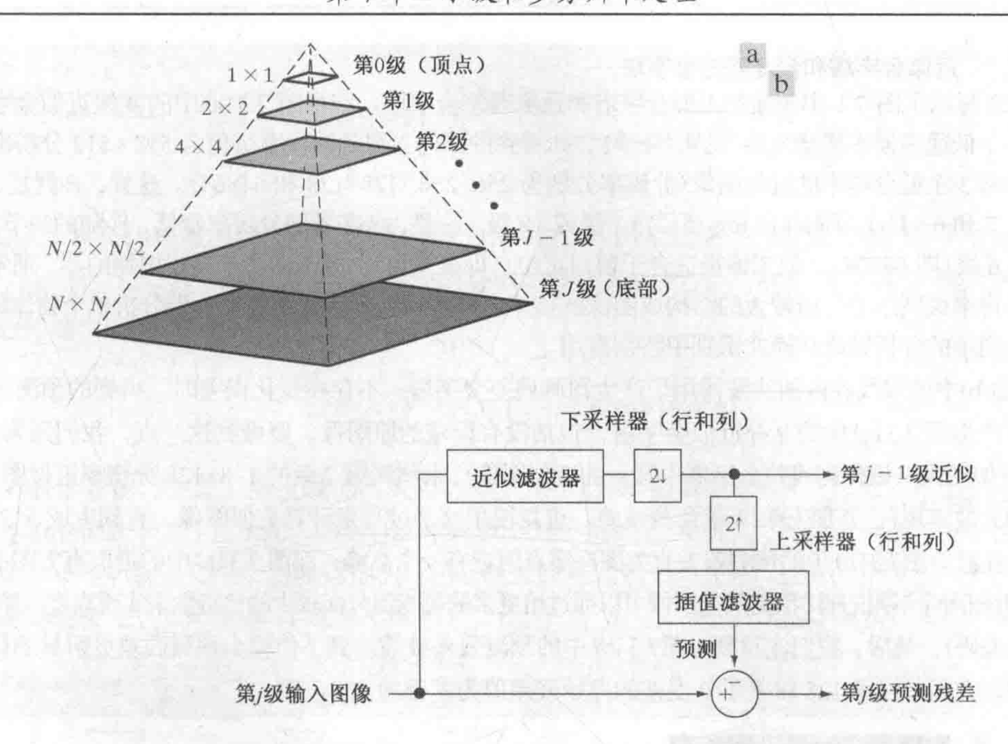


图 7.2　(a) 一个图像金字塔；(b) 创建近似和预测残差金字塔的一个简单系统

步骤 3　计算步骤 2 的预测图像和步骤 1 的输入之间的差。把得到的结果放在预测残差金字塔的第 j 级。

P 次迭代后的结论(即随后的迭代中 $j = J - P + 1$)表明，第 $J - P$ 级近似输出放在预测残差金字塔的第 $J - P$ 级处。如果不需要预测残差金字塔，则这些操作连同步骤 2、步骤 3、上采样器、内插滤波器以及图 7.2(b) 中的加法器都可以省略。

多种近似和内插滤波器可以合并到图 7.2(b) 所示的系统中。滤波通常在空间域执行(见 3.4 节)。有用的近似滤波技术包括：邻域平均(见 3.5.1 节)，它产生平均金字塔；低通高斯滤波(见 4.7.4 节和 4.8.3 节)，它产生高斯金字塔；不过滤，它产生子取样金字塔。2.4.4 节中描述的任何内插方法，包括最近邻、双线性和双三次内插，都可以合并到内插滤波器中。最后，我们注意到，图 7.2(b) 中的上采样和下采样方框用于加倍和减半计算得到的近似及预测图像的空间维数。给定一个整数变量 n 和一维取样序列 $f(n)$，则上采样序列 $f_{2\uparrow}(n)$ 如下式定义：

> 在本章中，我们将处理连续和离散函数以及连续和离散变量。除了二维图像 $f(x, y)$ 之外，其他以 $x, y, z, \cdots$ 表示的均是连续变量，而 $i, j, k, l, m, n, \cdots$ 则为离散变量。

$$f_{2\uparrow}(n) = \begin{cases} f(n/2), & n \text{ 为偶数} \\ 0, & \text{其他} \end{cases} \tag{7.1-1}$$

其中，如下标表示的那样，上采样以 2 为基数。基 2 下采样的互补操作定义为

$$f_{2\downarrow}(n) = f(2n) \tag{7.1-2}$$

上采样可视为在序列中的每个样本后插入 0；下采样可视为每隔一个样本就丢弃一个样本。图 7.2(b) 中分别以标为 2↑的上采样和标为 2↓的下采样方框来注释，指出二维输入的行和列，其中它们的操作就是上采样和下采样。与 4.11.1 节中可分的二维 DFT 一样，二维上采样和下采样可依次通过式(7.1-1)和式(7.1-2)定义的一维操作来执行。

486
~
487

例 7.1　近似金字塔和预测残差金字塔。

图 7.3 显示了图 7.1 中花瓶的近似金字塔和预测残差金字塔。产生图 7.3(a)中的 4 级近似金字塔时，使用了一个低通高斯平滑滤波器(见 4.7.4 节)。如看到的那样，得到的金字塔包含 512×512 分辨率的原图像(底部)和 3 个低分辨率的近似图像(分辨率分别为 256×256, 128×128 和 64×64)。这样，P 就是 3，出现级别 9, 8, 7 和 6，且不可能存在 $\log_2(512)+1$ 级或 10 级。注意，金字塔的分辨率越低，伴随的细节就越少。例如，第 6 级(即 64×64)近似图像很适合于窗口定位（即取景窗)，但不适合于寻找植物的茎。通常，金字塔的低分辨率级别用于分析较大的结构或图像的整体内容；而高分辨率图像适合于分析单个物体的特性。这种由粗到细的分析策略在模式识别中特别有用。

图 7.3(b)中的双线性内插滤波器用于产生预测残差金字塔。不存在量化误差时，得到的预测残差金字塔可用于产生图 7.3(a)中的互补近似金字塔，包括没有误差的原图像。要做到这一点，我们用第 6 级的 64×64 近似图像（该预测残差金字塔中唯一的近似图像)，来预测第 7 级的 128×128 分辨率近似图像(通过上采样和滤波实现)，并加上第 7 级预测残差。重复使用这个过程来计算近似图像，直到生成 512×512 的原图像。注意，图7.3(b)中的预测残差直方图在零点附近有一个高峰，而图 7.3(a)中的近似直方图中则没有高峰。与近似图像不同的是，预测残差图像可以通过给更多的可能值分配较少的比特数来实现高度压缩(见8.2.1节的变长编码)。最后，我们注意到，图7.3(b)中的预测残差被缩放到了使较小预测误差更明显的程度，但预测残差直方图是以第 128 级表示 0 误差的原始残差值为基础的。

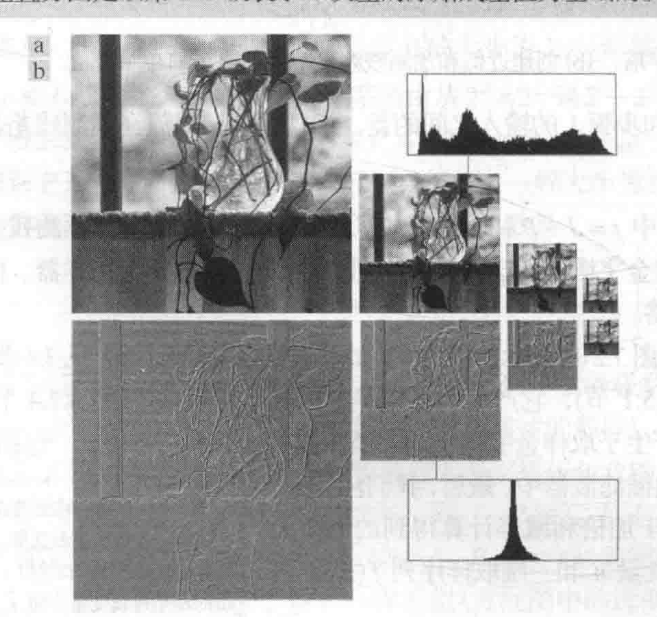


图 (a)中的近似金字塔称为高斯金字塔，因为构建该金字塔时使用了一个高斯滤波器。图(b)中的预测残差金字塔通常称为拉普拉斯金字塔；注意其与第 3 章中使用拉普拉斯算子滤波后得到的图像的相似性。

图 7.3　两种图像金字塔及它们的直方图：(a)近似金字塔；(b)预测残差金字塔

7.1.2　子带编码

另一种与多分辨率分析相关的重要图像处理技术是子带编码。在子带编码中，一幅图像被分解为一组频带受限的分量，称为子带。由于执行了分解，所以子带可以重组在一起，无误差地重建原始图像。因为分解和重建是借助于数字滤波器实现的，所以我们的讨论从数字信号处理(DSP)和数字信号滤波的简介开始。

术语“延迟”表明了一个基于时间的输入序列，并反映了数字信号滤波中输入通常是一个取样后的模拟信号这一事实。

考虑图7.4(a)中的简单数字滤波器，并注意到它是由三个基本部件组成的，即延迟单元、乘法器和加法器。沿这个滤波器的顶部，依次连接延迟单元，建立输入序列 $f(n)$的 $K-1$ 延迟(即右移)形式。例如，延迟序列 $f(n-2)$ 为

487 ~ 488

$$f(n-2)=\begin{cases}\vdots\\ f(0), & n=2\\ f(1), & n=2+1=3\\ \vdots\end{cases}$$

如图7.4(a)中灰色注释指出的那样，输入序列 $f(n)=f(n-0)$ 和延迟单元输出处的 $K-1$ 个延迟序列，表示为 $f(n-1), f(n-2),\cdots,f(n-K+1)$，它们分别与常数 $h(0),h(1),\cdots,h(K-1)$ 相乘，然后求和，可产生滤波后的输出序列：

$$\hat{f}(n)=\sum_{k=-\infty}^{\infty}h(k)f(n-k)=f(n)\star h(n) \tag{7.1-3}$$

式中，★ 表示卷积。注意，除变量的改变外，式(7.1-3)与第 4 章中定义离散卷积的公式(4.4-10)相同。图7.4(a)和式(7.1-3)中的乘数 K 称为滤波系数。每个系数定义一个滤波器抽头，它可视为计算式(7.1-3)中求和项所需要的分量，此时称滤波器为 K 阶的。

> 如果图 7.4(a)中滤波器的系数是使用0到 $K-1$ 之间的 n 个值来索引的，则式(7.1-3)中的求和限可简化为 0 到 $K-1$［像在式(4.4-10)中那样］。

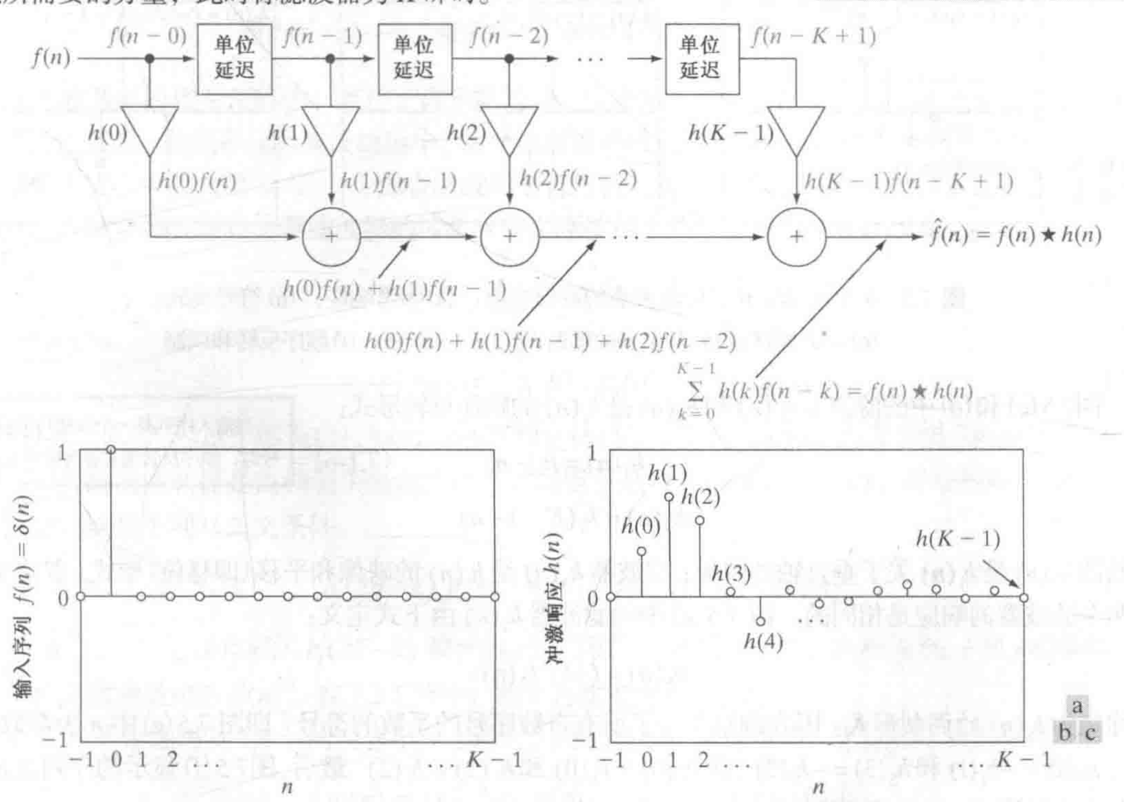


图 7.4　(a)数字滤波器；(b)单元离散冲激序列；(c)滤波器的冲激响应

如果图7.4(a)中滤波器的输入是图7.4(b)和 4.2.3 节中的离散单位冲激，则式(7.1-3)变为

$$\hat{f}(n)=\sum_{k=-\infty}^{\infty}h(k)\delta(n-k)=h(n) \tag{7.1-4}$$

也就是说，在式(7.1-3)中，将输入 $f(n)$ 用 $\delta(n)$ 来代替，并利用式(4.2-13)定义的离散单位冲激的取样特性，我们发现图 7.4(a)中滤波器的冲激响应是，定义该滤波器的滤波器系数的 K 元素序列。物理上，单位脉

> 在本章的剩余内容中，"滤波器 $h(n)$"将用于指其冲激响应为 $h(n)$的滤波器。

489~490

冲越过滤波器的顶端，从左到右移动(从一个单元延迟到下一个单元延迟)，产生一个输出，假设该输出就是延迟冲激位置处的系数值。因为有 K 个系数，所以冲激响应的长度为 K，且这种滤波器称为有限冲激响应滤波器(FIR)。

图 7.5 显示了 6 个功能上相关的滤波器的冲激响应。图7.5(b)中的 $h_2(n)$ 是图7.5(a)中 $h_1(n)$ 的符号反转形式(即关于水平轴反转)，即

$$h_2(n) = -h_1(n) \tag{7.1-5}$$

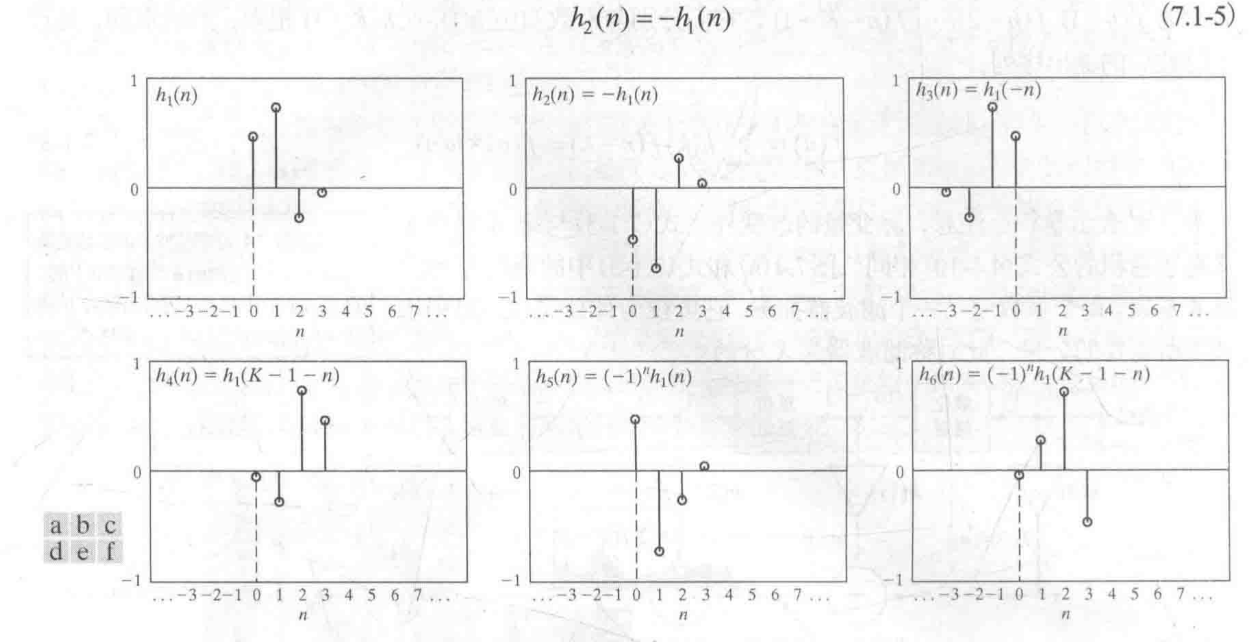


图 7.5 6 个功能上相关的滤波器的冲激响应：(a)参考响应；(b)符号反转；(c)~(d)顺序反转(与引入的延迟不同)；(e)调制；(f)顺序反转和调制

图7.5(c)和(d)中的滤波器 $h_3(n)$ 和 $h_4(n)$ 是 $h_1(n)$ 的顺序反转形式：

> 当输入序列是一个取样后的模拟信号时，顺序反转通常称为时间反转。

$$h_3(n) = h_1(-n) \tag{7.1-6}$$

$$h_4(n) = h_1(K-1-n) \tag{7.1-7}$$

滤波器 $h_3(n)$ 是 $h_1(n)$ 关于垂直轴的映像；滤波器 $h_4(n)$ 是 $h_1(n)$ 的映像和平移(即移位)形式。忽略平移，这两个滤波器的响应是相同的。图 7.5(e)中的滤波器 $h_5(n)$ 由下式定义：

$$h_5(n) = (-1)^n h_1(n) \tag{7.1-8}$$

它称之为 $h_1(n)$ 的调制形式。因为调制改变了所有奇数序号的系数的符号 [即图 7.5(e)中 n 为奇数的系数]，$h_5(1) = -h_1(1)$ 和 $h_5(3) = -h_1(3)$，而 $h_5(0) = h_1(0)$ 和 $h_5(2) = h_1(2)$。最后，图7.5(f)显示的序列是 $h_1(n)$ 的顺序反转形式，它也被调制：

$$h_6(n) = (-1)^n h_1(K-1-n) \tag{7.1-9}$$

包含这个序列的目的是为了说明这样一个事实，即在规定两个滤波器之间的关系时，符号反转、顺序反转和调制有时是合并在一起的。

> 滤波器组是两个或两个以上滤波器的集合。

简单介绍数字信号滤波后，我们考虑图7.6(a)中的一个两波段子带编码和解码系统。如图中指出的那样，该系统由两个滤波器组构成，每个滤波器组包含图 7.4(a)中所示的两个 FIR 滤波器。注意，4 个 FIR 滤波器中的每个，在图7.6(a)中都被表示为一个方框，方框内写出了每个滤波器的冲激响应(和卷积符号)。包含滤波器 $h_0(n)$ 和 $h_1(n)$ 的分析滤波器组用于把输入序

490 ~ 491

列分成两个半长序列 $f_{\text{lp}}(n)$ 和 $f_{\text{hp}}(n)$，表示输入的子带。注意，滤波器 $h_0(n)$ 和 $h_1(n)$ 是半波段滤波器，它们的理想传递特性 H_0 和 H_1 如图7.6(b)所示。滤波器 $h_0(n)$ 是一个低通滤波器，其输出即子带 $f_{\text{lp}}(n)$ 称为 $f(n)$ 的近似；滤波器 $h_1(n)$ 是一个高通滤波器，其输出即子带 $f_{\text{hp}}(n)$ 称为 $f(n)$的高频部分或细节部分。综合滤波器组 $g_0(n)$ 和 $g_1(n)$ 将 $f_{\text{lp}}(n)$ 和 $f_{\text{hp}}(n)$ 合并，产生 $\hat{f}(n)$。子带编码的目的是，选择 $h_0(n), h_1(n), g_0(n)$ 和 $g_1(n)$，以便 $\hat{f}(n) = f(n)$，即子带编码和解码系统的输入和输出是相同的。完成这一任务时，就可以说最终系统采用了完美重建滤波器。

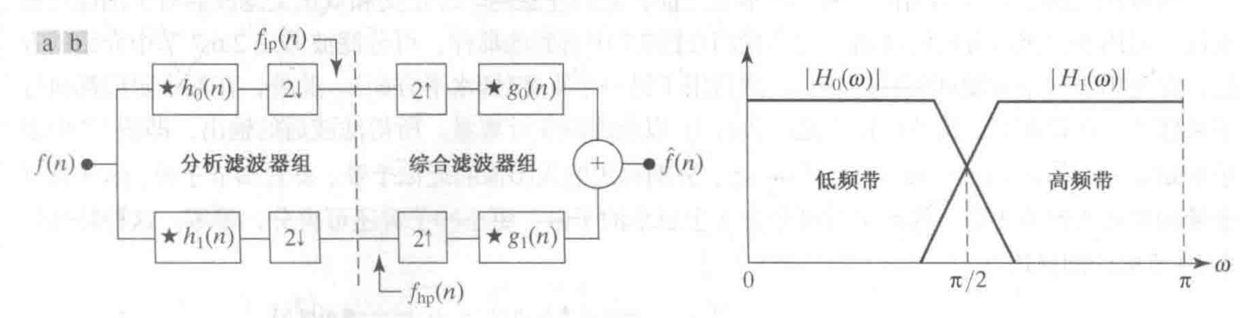


图 7.6　(a)一个二带子带编码和解码系统；(b)频谱可分性

在滤波器组的相关文献中，描述了许多两波段、实系数、FIR、完美重建滤波器组。在所有这些滤波器组中，综合滤波器都被调制为分析滤波器的形式——有（且只有）一个综合滤波器的符号被反转。要实现完美重构，综合滤波器和分析滤波器的冲激响应必须按如下两种方式之一关联起来：

> 式(7.1-10)至式(7.1-14)的详细描述见关于滤波器组的文献(例如，参见 Vetterli and Kovacevic[1995])。

$$g_0(n) = (-1)^n h_1(n)\ ,\ g_1(n) = (-1)^{n+1} h_0(n) \tag{7.1-10}$$

或

$$g_0(n) = (-1)^{n+1} h_1(n)\ ,\ g_1(n) = (-1)^n h_0(n) \tag{7.1-11}$$

式(7.1-10)和式(7.1-11)中的 $h_0(n), h_1(n), g_0(n)$ 和 $g_1(n)$ 被视为交叉调制，因为在图 7.6(a)所示的方框图中，斜对着的滤波器是由调制关联起来的［调制因子是 $-(-1)^n$ 或 $(-1)^{n+1}$ 时，符号相反］。此外，可以看出它们满足下列双正交条件：

$$\langle h_i(2n-k), g_j(k)\rangle = \delta(i-j)\delta(n),\quad i,j = \{0,\ 1\} \tag{7.1-12}$$

式中，$\langle h_i(2n-k), g_i(k)\rangle$ 表示 $h_i(2n-k)$ 和 $g_j(k)$ 的内积[①]。i 不等于 j 时，内积为 0；i 和 j 相等时，内积是单位离散冲激函数 $\delta(n)$。在 7.2.1 节中，将会再次考虑双正交性。

在推导子带编码及 7.4 节的快速小波变换时，特别重要的是滤波器，它排除了双正交性并要求

$$\langle g_i(n), g_j(n+2m)\rangle = \delta(i-j)\delta(m),\quad i,j = \{0,\ 1\} \tag{7.1-13}$$

上式定义了完美重建滤波器组的正交性。除式(7.1-13)外，可以证明正交滤波器满足如下两个条件：

$$\begin{aligned} g_1(n) &= (-1)^n g_0(K_{\text{even}} - 1 - n) \\ h_i(n) &= g_i(K_{\text{even}} - 1 - n),\quad i = \{0,\ 1\} \end{aligned} \tag{7.1-14}$$

式中，K_{even} 的下标用于指出滤波器系数的值必须能被 2 整除（即偶数）。如式(7.1-14)指出的那样，

① 序列 $f_1(n)$ 和 $f_2(n)$ 的向量内积是 $\langle f_1, f_2\rangle = \sum_n f_1^*(n) f_2(n)$，其中*表示复共轭操作。如果 $f_1(n)$ 和 $f_2(n)$ 是实的，则有 $\langle f_1, f_2\rangle = \langle f_2, f_1\rangle$。

491 ~ 493

综合滤波器 g_1 通过顺序反转和调制与 g_0 建立联系。另外，h_0 和 h_1 分别是综合滤波器 g_0 和 g_1 的顺序反转形式。这样，正交滤波器组就可围绕单一滤波器的冲激响应来开发，称之为原型；其余滤波器可以通过指定原型的冲激响应来计算。双正交滤波器组要求有两个原型；其余滤波器可通过式(7.1-10)和式(7.1-11)来计算。有用原型滤波器的产生，无论是归一化正交的还是双正交的，都超出了本章的讨论范围。这里我们只是简单地使用文献中已有的滤波器，为深入学习提供参考。

在使用二维子带编码的例子来总结本节之前，我们注意到一维正交和双正交滤波器对于图像处理来说，可用于二维可分的滤波器。就像我们在图7.7 中看到的那样，可分滤波器用 2.6.7 节中介绍的方法，首先用于一个维度(即垂直方向)，然后用于另一个维度(即水平方向)。此外，在两个阶段都执行下采样（一次是在第二次滤波操作之前执行），以全面减少计算量。所得滤波后的输出，即图7.7 中表示的 $a(m,n), d^V(m,n), d^H(m,n)$ 和 $d^D(m,n)$，分别称为输入图像的近似子带、垂直细节子带、水平细节子带和对角线细节子带。这些子带可分为 4 个更小的子带，更小的子带还可再分，等等。这种特性将在7.4 节更详细描述。

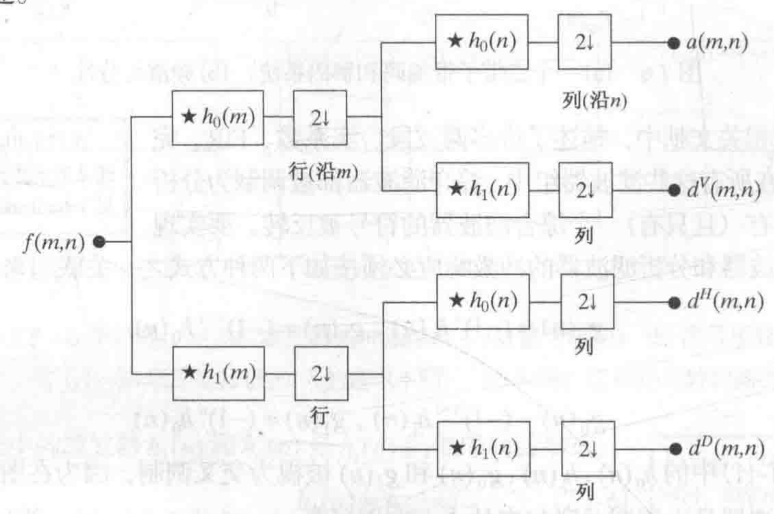


图 7.7 子带图像编码的一个二维 4 带宽滤波器组

例 7.2 图 7.1 中花瓶的一个 4 带宽子带编码。

图 7.8 显示了 4 个 8 抽头归一化正交滤波器的冲激响应。原型综合滤波器 $g_0(n)$，$0 \leqslant n \leqslant 7$ 的系数［图 7.8(c)］在表7.1 中已定义(Daubechies [1992])。其余归一化正交滤波器的系数可用式(7.1-14)算出。借助于图 7.5，注意(视觉观察)图 7.8 中分析和综合滤波器的交叉调制。以数字形式来说明滤波器是双正交的［满足式(7.1-12)］和归一化正交的［满足式(7.1-13)］相对比较容易。因此，图 7.8 中的 Daubechies 8 抽头滤波器支持分解输入的无误差重建。

表 7.1 $g_0(n)$ 的 Daubechies 8 抽头归一化正交滤波器系数(Daubechies [1992])

n	$g_0(n)$
0	0.23037781
1	0.71484657
2	0.63088076
3	−0.02798376
4	−0.18703481
5	0.03084138
6	0.03288301
7	−0.01059740

图7.1 中，512×512 花瓶图像以图 7.8 的滤波器为基础分离了 4 个子带，如图 7.9 所示。该图像的每个四分之一象限都是 256×256 大小的子带。从左上角开始并按顺时针方式，这 4 个象限分别包含近似子带 a、水平细节子带 d^H、对角线细节子带 d^D 和垂直细节子带 d^V。除了图 7.9(a)中的近似子带外，所有子带均已被缩放，以便它们的基本结构更加明显。注意，图7.9(b)和(c)的 d^H 子带和 d^V 子带中出现了可见的混淆效应。窗户区域的波状线是对图7.1 中一个刚好能分辨的窗口

493 ~ 494

下采样造成的。不管是否混淆，原始图像都可以由图 7.9 的子带无误差地重建。所需综合滤波器 $g_0(n)$ 和 $g_1(n)$ 由表7.1 和式(7.1-14)决定，并且合并成了一个滤波器组，它们可大致反映图7.7的系统。在新滤波器组中，滤波器 $h_i(n)$，$i = \{0, 1\}$被它们的 $g_i(n)$ 副本取代，并加上了取样器和求和器。

关于混淆效应的详细信息，请参阅 4.5.4 节。

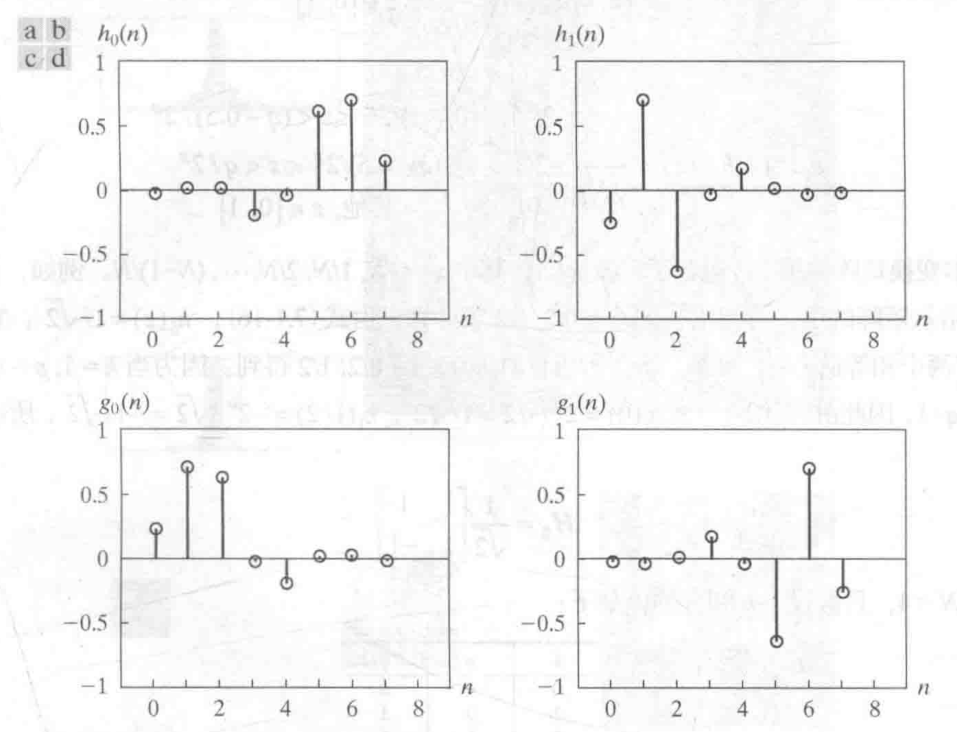


图 7.8　4 个 8 抽头 Daubechies 归一化正交滤波器的冲激响应。$g_0(n)$，$0 \leqslant n \leqslant 7$ 的值见表 7.1

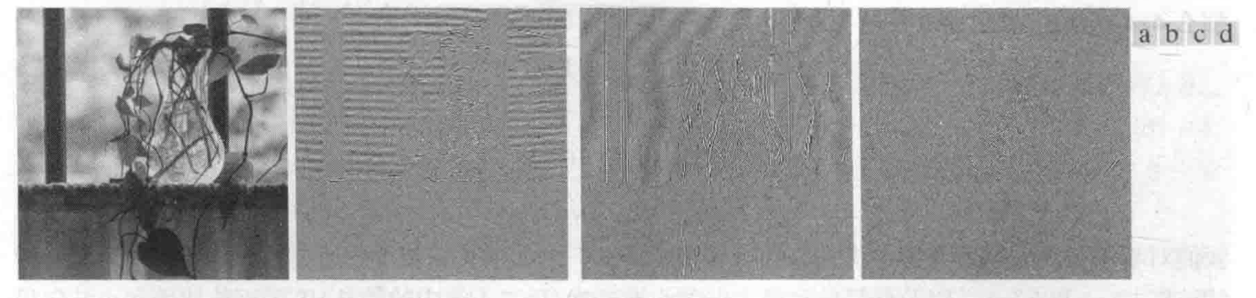


图 7.9　使用图 7.7 所示的子带编码系统对图 7.1 中的花瓶进行 4 子带分离。得到的 4 个子带分别是：(a)近似子带；(b)水平细节子带；(c)垂直细节子带；(d)对角线细节子带

7.1.3　哈尔变换

第三个也是最后一个与多分辨率分析紧密联系的图像操作是哈尔变换(Haar[1910])。在本章的内容中，哈尔变换的重要性体现在其基函数(在下面定义)是已知的最古老、也最简单的正交小波。它们将在本节下面的很多实例中得到应用。

以 2.6.7 节的讨论作为参考，哈尔变换可用如下矩阵形式表示：

$$\boldsymbol{T} = \boldsymbol{HFH}^{\mathrm{T}} \tag{7.1-15}$$

式中，$\boldsymbol{F}$ 是一个 $N \times N$ 图像矩阵，$\boldsymbol{H}$ 是一个 $N \times N$ 哈尔变换矩阵，$\boldsymbol{T}$ 是一个 $N \times N$ 变换结果。转置是必要的，因为 $\boldsymbol{H}$ 不对称；在 2.6.7 节的式(2.6-38)中，变换矩阵假定是对称的。对于哈尔变换，$\boldsymbol{H}$ 包含哈

495~496

尔基函数 $h_k(z)$，它们定义在连续闭区间 $z \in [0,1], k = 0,1,2,\cdots,N-1$ 上，其中 $N = 2^n$。要生成矩阵 $\boldsymbol{H}$，需要定义整数 k，即 $k = 2^p + q - 1$，其中 $0 \leqslant p \leqslant n-1$，当 $p=0$ 时，$q=0$ 或 1，而当 $p \neq 0$ 时，$1 \leqslant q \leqslant 2^p$。因此，哈尔基函数为

$$h_0(z) = h_{00}(z) = \frac{1}{\sqrt{N}}, \ z \in [0, 1] \tag{7.1-16}$$

和

$$h_k(z) = h_{pq}(z) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2}, & (q-1)/2^p \leqslant z < (q-0.5)/2^p \\ -2^{p/2}, & (q-0.5)/2^p \leqslant z < q/2^p \\ 0\ , & \text{其他}, \ z \in [0, 1] \end{cases} \tag{7.1-17}$$

$N \times N$ 哈尔变换矩阵的第 i 行包含了元素 $h_i(z)$，其中 $z = 0/N, 1/N, 2/N, \cdots, (N-1)/N$。例如，如果 $N=2$，那么 2×2 哈尔矩阵的第一行使用 $h_0(z), z = 0/2, 1/2$ 来计算。由式(7.1-16)，$h_0(z) = 1/\sqrt{2}$，它与 z 无关，因此 $\boldsymbol{H}_2$ 有两个相等的 $1/\sqrt{2}$ 元素。第二行由计算 $h_1(z), z = 0/2, 1/2$ 得到。因为当 $k=1, p=0$ 且 $q=1$ 时有 $k=2^p+q-1$，因此由式(7.1-17)有 $h_1(0) = 2^0/\sqrt{2} = 1/\sqrt{2}$，$h_1(1/2) = -2^0/\sqrt{2} = -1/\sqrt{2}$，所以 2×2 哈尔矩阵是

$$\boldsymbol{H}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{7.1-18}$$

如果 $N=4$，且假设 k, q 和 p 的值如下：

k	p	q
0	0	0
1	0	1
2	1	1
3	1	2

那么 4×4 变换矩阵 $\boldsymbol{H}_4$ 为

$$\boldsymbol{H}_4 = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \tag{7.1-19}$$

我们对哈尔变换的主要兴趣在于，$\boldsymbol{H}_2$ 的行可用于定义一个 2 抽头完美重建滤波器组(见前一节)的分析滤波器 $h_0(n)$ 和 $h_1(n)$，以及最简单且最古老的小波变换(见 7.4 节中的例 7.10)的缩放比例和小波向量(分别在 7.2.2 节和 7.2.3 节中定义)。与其用计算哈尔变换来结束本节，不如用一个例子来说明分解方法的影响，剩余章节将要描述的方法中已经考虑到这一点。

例 7.3　离散小波变换中的哈尔函数。

图 7.10(a)显示了图 7.1 中的 512×512 图像的分解，它结合了金字塔编码、子带编码和哈尔变换(到目前为止我们已探讨了这三种技术)的关键性质。称其为离散小波变换(在本章稍后阐述)的原因是，这种表示是由如下重要性质来表征的。

1. 除了图 7.10(a)左上角的子图像外，局部直方图非常相似。许多像素接近于零。因为这些子图像(左上角的子图像除外)已被缩放，以使得基本结构更为明显，所显示的直方图在灰度 128 处有峰值(零值已被标定为中等灰度)。分解中的大量零值使得图像对于压缩来说是优秀的候选者(见第 8 章)。

496 ~ 497

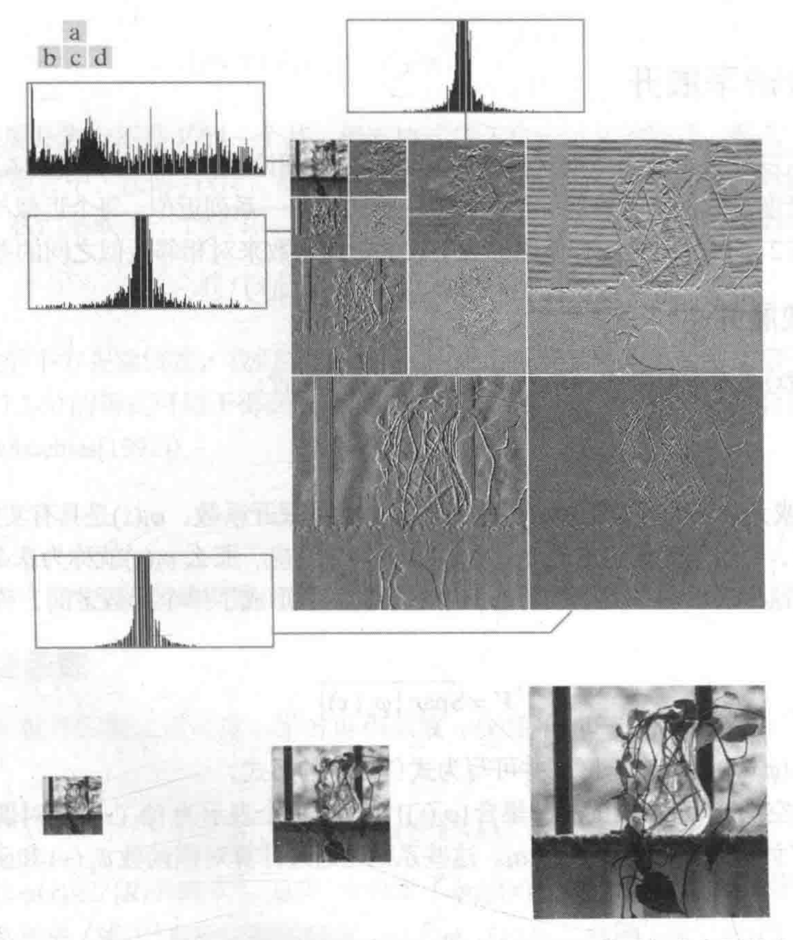


图 7.10　(a)用 $\boldsymbol{H}_2$ 哈尔基函数的离散小波变换，并显示了其局部直方图的变化；(b)~(d)由(a)得到的几种不同的近似(64 × 64, 128 × 128 和 256 × 256)

2. 采用类似于使用图 7.3(b)中的预测残差金字塔的各个级别来创建不同分辨率近似图像的方法，图 7.10(a)的子图像可用来构建图7.1 中原始花瓶图像的粗/细分辨率近似。图7.10(b)到(d)中大小分别为 64×64, 128×128 和 256×256 的图像是由图 7.10(a)中的子图像生成的。因此有可能完美重建 512×512 的原始图像。
3. 类似于图7.9 中的子带编码分解，使用图7.7 中给出的实系数FIR 滤波器组产生图 7.10(a)。类似于图 7.9 的一幅 4 子带图像生成后，256 × 256 的近似子带被分解并被 4 个 128 × 128 的子带代替(使用相同的滤波器组)，得到的近似子带被再次分解并被 4 个 64 × 64 的子带代替。这一处理产生了唯一的子图像排列，并表征了离散小波变换的特性。正如我们从图像的右下角移到左上角那样，图 7. 10(a)的子图像尺寸逐渐变小。
4. 图 7.10(a)不是图 7.1 中图像的哈尔变换。尽管用于产生这种分解的滤波器组系数取自哈尔变换矩阵 $\boldsymbol{H}_2$，但多种归一化正交和双正交滤波器组系数可在离散小波变换中使用。
5. 如 7.4 节中将说明的那样，图7.10(a)中的每幅子图像都描绘了原始图像中空间频率的一个特定频带。另外，许多子图像显示了方向敏感性[例如，图7.10(a)右上角的子图像获取原图像中的水平边缘信息]。

考虑到令人印象深刻的这些特性，图7.10(a)的离散小波变换可用共 4 个滤波器系数的 2 抽头数字滤波器来产生。

497~499

7.2 多分辨率展开

前一节介绍了三种著名的图像处理技术，它们在数学理论多分辨率分析(MRA)中扮演了重要角色。在MRA中，尺度函数被用于建立一个函数或一幅图像的一系列近似，每个近似与其最近邻近似在分辨率方面都用基2来区分。然后，使用称为小波的附加函数来对相邻近似之间的差进行编码。

7.2.1 级数展开

信号或函数$f(x)$通常能分解为一系列展开函数的线性组合：

$$f(x)=\sum_{k}\alpha_{k}\varphi_{k}(x) \tag{7.2-1}$$

式中，k是有限和或无限和的整数下标，α_k是具有实数值的展开系数，$\varphi_k(x)$是具有实数值的展开函数。如果展开是唯一的，即对于任何给定的c只有一组α_k与之对应，那么$\varphi_k(x)$就称为*基函数*，并且展开集合$\{\varphi_k(x)\}$就称为可这样表示的一类函数的基。可展开的函数形成了一个函数空间，称为展开集合的闭合跨度，表示为

$$V=\overline{\underset{k}{\text{Span}}\{\varphi_{k}(x)\}} \tag{7.2-2}$$

$f(x)\in V$指$f(x)$属于$\{\varphi_k(x)\}$的闭合跨度，并可写为式(7.2-1)的形式。

对于任意函数空间V及相应的展开集合$\{\varphi_k(x)\}$，都有一个表示为$\{\tilde{\varphi}_k(x)\}$的对偶函数集合，它可用于对任意$f(x)\in V$计算式(7.2-1)的系数α_k。这些系数是通过计算对偶函数$\tilde{\varphi}_k(x)$和函数$f(x)$的内积[①]得到的，即

$$\alpha_{k}=\langle\tilde{\varphi}_{k}(x),f(x)\rangle=\int\tilde{\varphi}_{k}^{*}(x)f(x)\mathrm{d}x \tag{7.2-3}$$

式中，*表示复共轭操作。依靠展开集合的正交性，该计算假定是三种可能形式中的一种。本章末尾的习题7.10用二维欧氏空间中的向量说明了这三种情况。

情况1：如果展开函数构成了V的一个正交基，即

$$\langle\varphi_{j}(x),\varphi_{k}(x)\rangle=\delta_{jk}=\begin{cases}0, & j\neq k\\ 1, & j=k\end{cases} \tag{7.2-4}$$

则该基与它的对偶基相等，即$\varphi_k(x)=\tilde{\varphi}_k(x)$，式(7.2-3)变成

$$\alpha_{k}=\langle\varphi_{k}(x),f(x)\rangle \tag{7.2-5}$$

α_k由基函数与$f(x)$的内积来计算。

情况2：如果展开函数本身不正交，但却是V的一个正交基，那么

$$\langle\varphi_{j}(x),\varphi_{k}(x)\rangle=0,\ j\neq k \tag{7.2-6}$$

且基函数及其对偶称为*双正交函数*。用式(7.2-3)来计算α_k，双正交基及其对偶有下列形式：

---

499 ~ 500

① 两个实或复值函数$f(x)$和$g(x)$的内积是$\langle f(x),g(x)\rangle=\int f^{*}(x)g(x)\mathrm{d}x$。如果$f(x)$是实的，则$f^{*}(x)=f(x)$且$\langle f(x),g(x)\rangle=\int f(x)g(x)\mathrm{d}x$。

$$\langle \varphi_j(x), \tilde{\varphi}_k(x) \rangle = \delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \tag{7.2-7}$$

情况 3：如果展开集合不是 V 的一个基，但支持式(7.2-1)中定义的展开，那么它就是一个跨度集合，在该跨度集合中，任何 $f(x) \in V$ 都有一个以上 α_k 的集合。展开函数及其对偶可以说是超完备的或冗余的。它们形成了一个框架，其中，对于某些 $A > 0$，$B < \infty$ 及所有 $f(x) \in V$，有[①]

$$A\|f(x)\|^2 \leqslant \sum_k \left|\langle \varphi_k(x), f(x) \rangle\right|^2 \leqslant B\|f(x)\|^2 \tag{7.2-8}$$

用 $f(x)$ 的范数的平方去除该式，我们看到 A 和 B 构成了扩展系数与该函数的归一化内积。类似式(7.2-3)和式(7.2-5)的等式可用于得到框架的展开系数。若 $A = B$，则展开集合称为紧框架，并且可以证明(Daubechies[1992])

$$f(x) = \frac{1}{A}\sum_k \langle \varphi_k(x), f(x) \rangle \varphi_k(x) \tag{7.2-9}$$

除 A^{-1} 项之外，它是框架冗余的度量，这与将式(7.2-5)(对于正交基)代入式(7.2-1)得到的表达式相等。

7.2.2　尺度函数

考虑由整数平移和实数二值尺度、平方可积函数 $\varphi(x)$ 组成的展开函数集合，即集合 $\{\varphi_{j,k}(x)\}$，其中

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) \tag{7.2-10}$$

对所有的 $j, k \in \boldsymbol{Z}$ 和 $\varphi(x) \in L^2(\boldsymbol{R})$ 都成立[②]。这里，k 决定了 $\varphi_{j,k}(x)$ 沿 x 轴的位置，j 决定了 $\varphi_{j,k}(x)$ 的宽度，即它沿 x 轴是宽还是窄。项 $2^{j/2}$ 控制函数的幅度。由于 $\varphi_{j,k}(x)$ 的形状随 j 发生变化，所以 $\varphi(x)$ 称为尺度函数。适当选择 $\varphi(x)$，可使 $\{\varphi_{j,k}(x)\}$ 跨越 $L^2(\boldsymbol{R})$，其中后者是所有可度量的、平方可积函数的集合。

若将式(7.2-10)中的 j 限制为某个指定值，如 $j = j_0$，则得到的展开集合 $\{\varphi_{j0,k}(x)\}$ 是 $\{\varphi_{j,k}(x)\}$ 的一个子集，它跨越 $L^2(\boldsymbol{R})$ 的子空间。使用前一节中的符号，我们可将该子空间定义为

$$V_{j0} = \overline{\underset{k}{\text{Span}}\{\varphi_{j0,k}(x)\}} \tag{7.2-11}$$

即 V_{j0} 是 $\varphi_{j0,k}(x)$ 在 k 上的一个跨度。$f(x) \in V_{j0}$ 时，我们可写出

$$f(x) = \sum_k \alpha_k \varphi_{j0,k}(x) \tag{7.2-12}$$

更一般地，对于任何 j，我们将 k 上跨越的子空间表示为

$$V_j = \overline{\underset{k}{\text{Span}}\{\varphi_{j,k}(x)\}} \tag{7.2-13}$$

如下例所示，增加 j 就会增加 V_j 的大小，进而允许子空间中包含具有较小变化的变量或较细的细节函数，原因如下：随着 j 的增大，用于表示子空间函数的 $\varphi_{j,k}(x)$ 会变得较窄，且 x 有较小变化就可分开。

①表示为 $\|f(x)\|$ 的 $f(x)$ 的范数定义为 $f(x)$ 与自身内积的绝对值的平方根。

② 符号 $L^2(\boldsymbol{R})$ 中，$\boldsymbol{R}$ 是实数集，表示可度量的、平方可积的一维函数集合，$\boldsymbol{Z}$ 是整数集。

500 ~ 501

例 7.4　哈尔尺度函数

考虑单位高度、单位宽度的尺度函数(Haar[1910])

$$\varphi(x)=\begin{cases}1, & 0\leqslant x<1\\ 0, & \text{其他}\end{cases} \tag{7.2-14}$$

图7.11(a)到图7.11(d)显示了多个展开函数中的4个函数，这些展开函数是通过将脉冲型尺度函数代入式(7.2-10)中得到的。注意，$j=1$时，图7.11(c)和(d)中的展开函数是图7.11(a)和(b)中$j=0$时展开函数的宽度的一半。对于x的一个给定区间，我们可以按V_1的尺度函数的两倍定义V_0的尺度函数(例如，对于区间$0\leqslant x<1$，V_1的$\varphi_{1,0}$和$\varphi_{1,1}$相对于V_0的$\varphi_{0,0}$)。

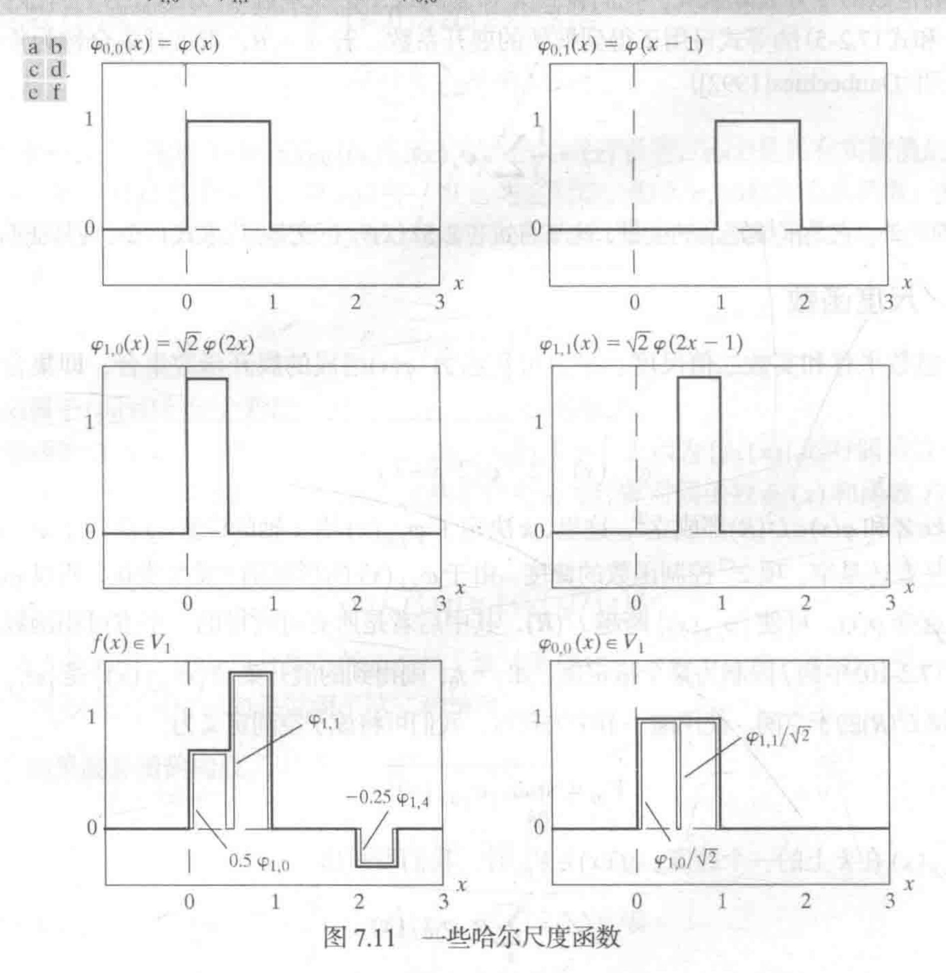


图 7.11　一些哈尔尺度函数

图7.11(e)显示了子空间V_1中的一个成员。该函数不属于V_0，因为图7.11(a)和图7.11(b)中的V_0展开函数太粗糙而无法表示它。我们需要如图7.11(c)和图7.11(d)所示的高分辨率函数。如图7.11(e)所示，该函数可用三项展开来表示：

$$f(x)=0.5\varphi_{1,0}(x)+\varphi_{1,1}(x)-0.25\varphi_{1,4}(x)$$

为总结该例子，图7.11(f)表明将$\varphi_{0,0}(x)$的分解作为V_1展开函数的和。采用类似的方式，V_0的展开函数可以使用下式分解：

$$\varphi_{0,k}(x)=\frac{1}{\sqrt{2}}\varphi_{1,2k}(x)+\frac{1}{\sqrt{2}}\varphi_{1,2k+1}(x)$$

502 ~ 503

因此，如果$f(x)$是V_0的元素，那么它也是V_1的元素，因为所有的V_0展开函数都包含在V_1中。数学上，我们可以写出V_0是V_1的一个子空间，记为$V_0\subset V_1$。

在上例中，简单尺度函数满足多分辨率分析的 4 个基本要求(Mallat[1989a])：

MRA 要求 1：尺度函数对其整数平移是正交的。

很容易看出，在哈尔函数的情况下，无论什么时候，只要尺度函数的值是 1，其整数平移就是 0，所以二者的乘积为 0。哈尔尺度函数是紧支撑的，即除了称为支撑区的有限区间外，函数值都为 0。事实上，支撑的宽度是 1；半开区间[0, 1)外其值为 0。注意，尺度函数的支撑宽度大于 1 时，正交整数平移的要求将很难满足。

MRA 要求 2：低尺度的尺度函数跨越的子空间，嵌套在高尺度跨越的子空间内。

如图 7.12 所示，包含高分辨率函数的子空间必须同时包含所有的低分辨率函数。也就是说，

$$V_{-\infty} \subset \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_{\infty} \qquad (7.2\text{-}15)$$

此外，这些子空间还满足直观条件，即如果 $f(x) \in V_j$，那么 $f(2x) \in V_{j+1}$。哈尔尺度函数满足该要求的事实，并不意味着任何支撑宽度为 1 的函数都自动满足该条件。作为练习，请读者证明同样简单的函数

$$\varphi(x) = \begin{cases} 1, & 0.25 \leqslant x < 0.75 \\ 0, & \text{其他} \end{cases}$$

不是用于多分辨率分析的一个有效尺度函数(见习题 7.11)。

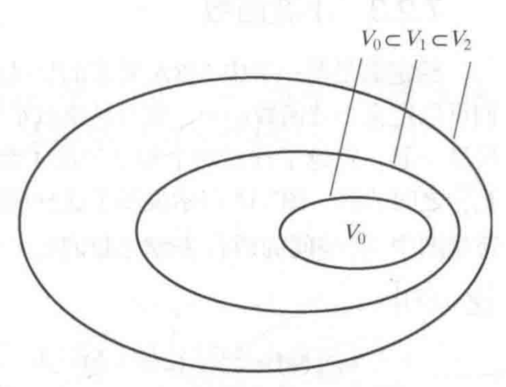


图 7.12　尺度函数跨越的嵌套函数空间

MRA 要求 3：唯一对所有 V_j 通用的函数是 $f(x) = 0$。

若考虑可能最粗糙的展开函数(即 $j = -\infty$)，则唯一的表达函数就是没有信息的函数，即

$$V_{-\infty} = \{0\} \qquad (7.2\text{-}16)$$

MRA 要求 4：任何函数都可以按任意精度表示。虽然不可能以任意粗糙的分辨率来展开一个特定的 $f(x)$，例如图 7.11(e)所示函数的情况，但所有可度量的、平方可积的函数都可以由尺度函数在 $j \to \infty$ 的限制下表示，即

$$V_{\infty} = \left\{L^2(\boldsymbol{R})\right\} \qquad (7.2\text{-}17)$$

在这些条件下，子空间 V_j 的展开函数可以表述为子空间 V_{j+1} 的展开函数的加权和。使用式(7.2-12)，令

$$\varphi_{j,k}(x) = \sum_n \alpha_n \varphi_{j+1,n}(x)$$

式中，为清楚起见，求和的下标改写成了 n。将来自式(7.2-10)的 $\varphi_{j+1,n}(x)$代入上式，并将变量 α_n 改写成 $h_\varphi(n)$，上式变成

$$\varphi_{j,k}(x) = \sum_n h_\varphi(n) 2^{(j+1)/2} \varphi(2^{j+1}x - n)$$

因为 $\varphi(x) = \varphi_{0,0}(x)$，$j$ 和 k 都可以置为 0，以得到较简单的无下标表达式

$$\varphi(x) = \sum_n h_\varphi(n) \sqrt{2} \varphi(2x - n) \qquad (7.2\text{-}18)$$

该递归等式中的系数 $h_\varphi(n)$称为尺度函数系数；h_φ 为尺度向量。式(7.2-18)是多分辨率分析的基础，称为改进等式、MRA 等式或扩张等式。它表明任意子空间的展开函数，都可以由其本身的双倍分辨率副本建立，即来自相邻较高分辨率空间的展开函数。参考子空间 V_0 的选择是任意的。

503
~
504

例 7.5 哈尔尺度函数系数。

式(7.2-14)所示哈尔函数的尺度函数系数是 $h_\varphi(0)=h_\varphi(1)=1/\sqrt{2}$，即式(7.1-18)中矩阵 $\boldsymbol{H}_2$ 的第一行。因此，由式(7.2-18)可得

$$\varphi(x)=\frac{1}{\sqrt{2}}\left[\sqrt{2}\varphi(2x)\right]+\frac{1}{\sqrt{2}}\left[\sqrt{2}\varphi(2x-1)\right]$$

图 7.11(f)以图形方式说明了 $\varphi_{0,0}(x)$的分解，其中前面表达式中方括号内的项被视为 $\varphi_{1,0}(x)$和 $\varphi_{1,1}(x)$。化简后得到 $\varphi(x)=\varphi(2x)+\varphi(2x-1)$。

7.2.3 小波函数

给定满足前一节中 MRA 要求的尺度函数，我们可以定义小波函数$\psi(x)$，它与其整数平移及二值尺度一起，跨越了任意两个相邻尺度子空间 V_j 和 V_{j+1}之间的差。图7.13 图示说明了这种情况。对于跨越图中 W_j 空间的所有 $k\in\boldsymbol{Z}$，我们定义小波集合 $\left\{\psi_{j,k}(x)\right\}$，

$$\psi_{j,k}(x)=2^{j/2}\psi(2^j x-k) \tag{7.2-19}$$

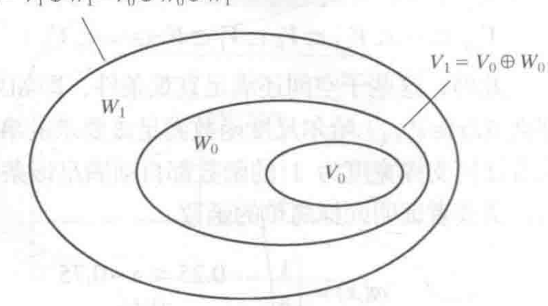


图 7.13 尺度函数与小波函数空间之间的关系

使用尺度函数，可以写出

$$W_j=\overline{\operatorname*{Span}_k\left\{\psi_{j,k}(x)\right\}} \tag{7.2-20}$$

并注意到如果$f(x)\in W_j$，那么有

$$f(x)=\sum_k \alpha_k\psi_{j,k}(x) \tag{7.2-21}$$

图 7.13 中的尺度函数和小波函数子空间由下式联系起来：

$$V_{j+1}=V_j\oplus W_j \tag{7.2-22}$$

式中，$\oplus$ 表示空间的并集(类似于集合的并集)。V_{j+1} 中 V_j的正交补集是 W_j，且 V_j中的所有成员与 W_j 中的所有成员都正交。因此，

$$\left\langle\varphi_{j,k}(x),\psi_{j,l}(x)\right\rangle=0 \tag{7.2-23}$$

对所有适当的$j,k,l\in\boldsymbol{Z}$都成立。

现在我们可以将所有可度量的、平方可积的函数空间表示为

$$L^2(\boldsymbol{R})=V_0\oplus W_0\oplus W_1\oplus\cdots \tag{7.2-24}$$

或

$$L^2(\boldsymbol{R})=V_1\oplus W_1\oplus W_2\oplus\cdots \tag{7.2-25}$$

甚至是

505~506

$$L^2(\boldsymbol{R})=\cdots\oplus W_{-2}\oplus W_{-1}\oplus W_0\oplus W_1\oplus W_2\oplus\cdots \tag{7.2-26}$$

上式中消除了尺度函数，并且函数仅用小波项来表示［即在式(7.2-26)中仅有小波函数空间］。注意，

若 $f(x)$ 是 V_1 而非 V_0 的元素，则使用式(7.2-24)的展开包含使用 V_0 尺度函数的 $f(x)$ 的近似；来自 W_0 的小波将对这种近似与实际函数之间的差进行编码。由式(7.2-24)到式(7.2-26)，可推广得到

$$L^2(\boldsymbol{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots \tag{7.2-27}$$

式中，j_0 是任意开始尺度。

因为小波空间存在于由相邻较高分辨率尺度函数跨越的空间中(见图 7.13)，所以任何小波函数，类似式(7.2-18)中其尺度函数的对应部分，可以表示成平移后的双倍分辨率尺度函数的加权和。也就是说，可以写成

$$\psi(x) = \sum_n h_\psi(n)\sqrt{2}\varphi(2x-n) \tag{7.2-28}$$

式中，$h_\psi(n)$ 称为小波函数系数，h_ψ 为小波向量。利用图 7.13 中小波跨越正交补集空间和整数小波平移是正交的条件，可以证明 $h_\psi(n)$ 和 $h_\varphi(n)$ 按下述方式相关(例如，见 Burrus, Gopinath and Guo[1998])：

$$h_\psi(n) = (-1)^n h_\varphi(1-n) \tag{7.2-29}$$

注意该结果与式(7.1-14)的相似性，这一关系决定了正交子带编码和解码滤波器的冲激响应。

例 7.6　哈尔小波函数系数。

在上例中，哈尔尺度向量定义为 $h_\varphi(0) = h_\varphi(1) = 1/\sqrt{2}$。使用式(7.2-29)，相应的小波向量为 $h_\psi(0) = (-1)^0 h_\varphi(1-0) = 1/\sqrt{2}$ 和 $h_\psi(1) = (-1)^1 h_\varphi(1-1) = -1/\sqrt{2}$。注意，这些系数与式(7.1-18)中矩阵 $\boldsymbol{H}_2$ 的第二行对应。将这些值代入式(7.2-28)，得到 $\psi(x) = \varphi(2x) - \varphi(2x-1)$，其图形如图 7.14(a)所示。因此，哈尔小波函数为

$$\psi(x) = \begin{cases} 1, & 0 \leqslant x < 0.5 \\ -1, & 0.5 \leqslant x < 1 \\ 0, & \text{其他} \end{cases} \tag{7.2-30}$$

使用式(7.2-19)，我们现在可产生缩放且平移后的哈尔小波通式。图 7.14(b)和(c)中分别画出了两个这样的小波 $\psi_{0,2}(x)$ 和 $\psi_{1,0}(x)$。注意，空间 W_1 中的小波 $\psi_{1,0}(x)$ 要比空间 W_0 中的小波 $\psi_{0,2}(x)$ 窄；它可以用来表示更细微的细节。

图 7.14(d)显示了在子空间 V_1 而不在 V_0 中的函数。该函数在前述例子中曾考虑过[见图 7.11(e)]。虽然该函数不能在 V_0 中精确地表示，但式(7.2-22)指出它可以使用 V_0 和 W_0 的展开函数来展开。展开结果如下：

$$f(x) = f_a(x) + f_d(x)$$

式中，

$$f_a(x) = \frac{3\sqrt{2}}{4}\varphi_{0,0}(x) - \frac{\sqrt{2}}{8}\varphi_{0,2}(x) \quad 和 \quad f_d(x) = \frac{-\sqrt{2}}{4}\psi_{0,0}(x) - \frac{\sqrt{2}}{8}\psi_{0,2}(x)$$

式中，$f_a(x)$ 是使用 V_0 尺度函数对 $f(x)$ 的近似，而 $f_d(x)$ 为 $f(x) - f_a(x)$，用 W_0 小波的和来表示。这两个展开式如图 7.14(e)和(f)所示，将 $f(x)$ 用类似图 7.6 讨论过的高通和低通滤波器方法分成两部分。$f(x)$ 的低频部分在 $f_a(x)$ 中得到，它假定为每个整数区间上的平均值，而高频细节则在 $f_d(x)$ 中编码。

506
~
508

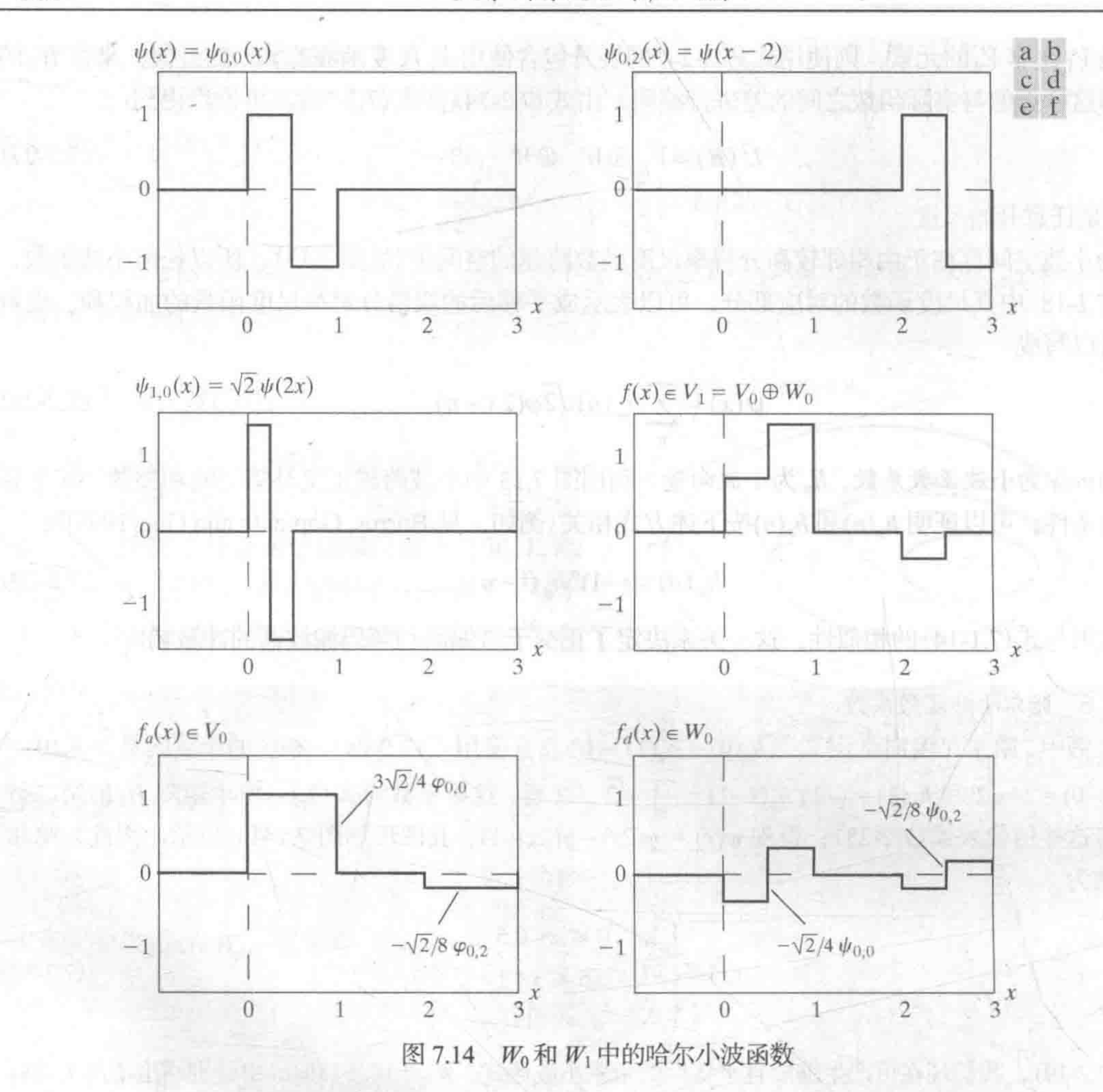


图 7.14　W_0 和 W_1 中的哈尔小波函数

7.3　一维小波变换

现在我们可以正式地定义几个紧密相关的小波变换：一般的小波级数展开、离散小波变换和连续小波变换。它们在傅里叶域中的对应形式分别是傅里叶级数展开、离散傅里叶变换和积分傅里叶变换。7.4 节将说明一种计算效率很高的称为快速小波变换的离散小波变换。

7.3.1　小波级数展开

首先定义与小波 $\psi(x)$ 和尺度函数 $\varphi(x)$ 相关的函数 $f(x)\in L^2(\boldsymbol{R})$ 的小波级数展开。根据式(7.2-27)，$f(x)$ 可表示为子空间 V_{j0} 中的尺度函数展开[式(7.2-12)定义了这样一个展开)和子空间 $W_{j0}, W_{j0+1},\cdots$ 中的某些小波函数[就像式(7.2-21)中定义的那样]，即

$$f(x)=\sum_k c_{j0}(k)\varphi_{j0,k}(x)+\sum_{j=j_0}^{\infty}\sum_k d_j(k)\psi_{j,k}(x) \tag{7.3-1}$$

其中，j_0 是任意的开始尺度，$c_{j0}(k)$ 和 $d_j(k)$ 分别是式(7.2-12)和式(7.2-21)中 α_k 的改写形式。$c_{j0}(k)$ 通常称为近似和/或尺度系数，$d_j(k)$ 称为细节和/或小波系数。这是因为式(7.3-1)中的第一个和式使用

尺度函数提供了 $f(x)$在尺度 j_0 处的近似［除非 $f(x)\in V_{j0}$ 使得尺度函数的和等于 $f(x)$］。对于第二个和式中的每个较高尺度 $j\geqslant j_0$，近似中添加了更细分辨率的函数——即小波的和，以便增加细节。如果展开函数形成了一个正交基或紧框架（通常情况下如此），那么基于式(7.2-5)和式(7.2-9)的展开系数计算如下：

$$c_{j0}(k)=\left\langle f(x),\varphi_{j0,k}(x)\right\rangle=\int f(x)\varphi_{j0,k}(x)\mathrm{d}x \tag{7.3-2}$$

> 因为 f 是实函数，故式(7.3-2)和式(7.3-3)的内积中不需要共轭。

和

$$d_j(k)=\left\langle f(x),\psi_{j,k}(x)\right\rangle=\int f(x)\psi_{j,k}(x)\mathrm{d}x \tag{7.3-3}$$

在式(7.2-5)和式(7.2-9)中，展开系数(即 α_k)定义为展开函数和用于展开的函数的内积。在式(7.3-2)和(7.3-3)中，展开函数是 $\varphi_{j0,k}$ 和 $\psi_{j,k}$，展开系数是 c_{j0} 和 d_j。如果展开函数是双正交基的一部分，那么这些等式中的 φ 项和 ψ 项必须用它们的对偶函数 $\tilde{\varphi}$ 和 $\tilde{\psi}$ 代替。

例 7.7　$y=x^2$ 的哈尔小波级数展开。

考虑图 7.15(a)所示的简单函数

$$y=\begin{cases}x^2, & 0\leqslant x\leqslant 1\\ 0, & \text{其他}\end{cases}$$

利用哈尔小波［见式(7.2-14)和式(7.2-30)］和一个开始尺度 $j_0=0$，可以用式(7.3-2)和式(7.3-3)计算出下述展开系数：

$$c_0(0)=\int_0^1 x^2\varphi_{0,0}(x)\,\mathrm{d}x=\int_0^1 x^2\mathrm{d}x=\left.\frac{x^3}{3}\right|_0^1=\frac{1}{3}$$

$$d_0(0)=\int_0^1 x^2\psi_{0,0}(x)\,\mathrm{d}x=\int_0^{0.5} x^2\mathrm{d}x-\int_{0.5}^1 x^2\mathrm{d}x=-\frac{1}{4}$$

$$d_1(0)=\int_0^1 x^2\psi_{1,0}(x)\,\mathrm{d}x=\int_0^{0.25} x^2\sqrt{2}\mathrm{d}x-\int_{0.25}^{0.5} x^2\sqrt{2}\mathrm{d}x=-\frac{\sqrt{2}}{32}$$

$$d_1(1)=\int_0^1 x^2\psi_{1,1}(x)\,\mathrm{d}x=\int_{0.5}^{0.75} x^2\sqrt{2}\mathrm{d}x-\int_{0.75}^{1} x^2\sqrt{2}\mathrm{d}x=-\frac{3\sqrt{2}}{32}$$

将这些值代入式(7.3-1)，得到如下小波级数展开：

$$y=\underbrace{\underbrace{\underbrace{\frac{1}{3}\varphi_{0,0}(x)}_{V_0}+\underbrace{\left[-\frac{1}{4}\psi_{0,0}(x)\right]}_{W_0}}_{V_1=V_0\oplus W_0}+\underbrace{\left[-\frac{\sqrt{2}}{32}\psi_{1,0}(x)-\frac{3\sqrt{2}}{32}\psi_{1,1}(x)\right]}_{W_1}}_{V_2=V_1\oplus W_1=V_0\oplus W_0\oplus W_1}+\cdots$$

上述展开中的第一项用 $c_0(0)$生成被展开函数的 V_0 子空间的近似。该近似如图7.15(b)所示，并且它是原始函数的平均值。第二项使用 $d_0(0)$通过从 W_0 子空间添加一级细节来改进该近似。添加的细节和得到的 V_1 近似分别如图 7.15(c)和图 7.15(d)所示。其他级别的细节由子空间 W_1 的系数 $d_1(0)$和 $d_1(1)$添加。该附加细节如图7.15(e)所示，得到的 V_2 近似如图7.15(f)所示。注意，展开函数现在开始接近原始函数。随着更高尺度(或更大级别的细节)的叠加，近似变得更接近函数的精确表示，在极限情况下，即当 $j\to\infty$ 时，可实现精确表示。

508～510

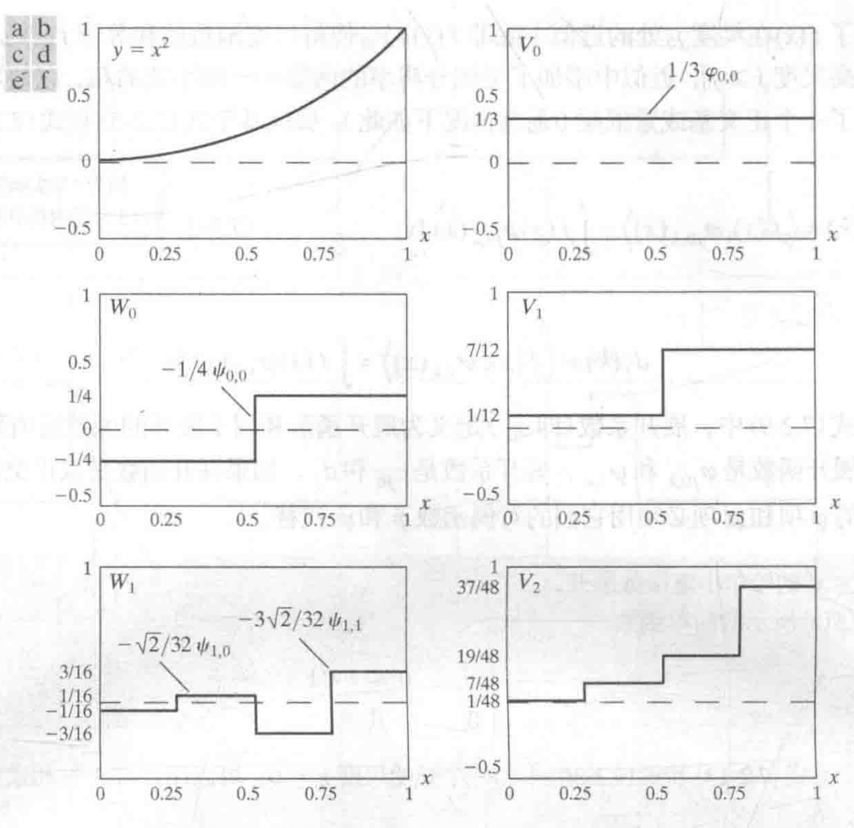


图 7.15　使用哈尔小波 $y=x^2$ 的小波级数展开

7.3.2　离散小波变换

与傅里叶级数展开相似，前一节的小波级数展开将一个连续变量函数映射为一系列系数。如果待展开的函数是离散的(即数字序列)，得到的系数就称之为离散小波变换(DWT)。例如，对于某些 $x_0, \Delta x$ 和 $n=0,1,2,\cdots,M-1$，若 $f(n)=f(x_0+n\Delta x)$，对 $f(x)$的小波级数展开系数[由式(7.3-2)和式(7.3-3)定义]就变成序列 $f(n)$的正向 DWT 系数：

$$W_\varphi(j_0,k)=\frac{1}{\sqrt{M}}\sum_n f(n)\varphi_{j_0,k}(n) \tag{7.3-5}$$

$$W_\psi(j,k)=\frac{1}{\sqrt{M}}\sum_n f(n)\psi_{j,k}(n), \qquad j\geqslant j_0 \tag{7.3-6}$$

在这些等式中，$\varphi_{j0,k}(n)$ 和 $\psi_{j,k}(n)$ 是基函数 $\varphi_{j0,k}(x)$ 和 $\psi_{j,k}(x)$ 的取样形式。例如，对于某些 $x_s, \Delta x_s$ 和 $n=0,1,2,\cdots,M-1$，有 $\varphi_{j0,k}(n)=\varphi_{j0,k}(x_s+n\Delta x_s)$。这样，我们就在基函数的支撑上使用 M 个等间隔的取样(见下面的例 7.8)。根据式(7.1-3)，反向 DWT 是

$$f(n)=\frac{1}{\sqrt{M}}\sum_k W_\varphi(j_0,k)\varphi_{j_0,k}(n)+\frac{1}{\sqrt{M}}\sum_{j=j_0}^{\infty}\sum_k W_\psi(j,k)\psi_{j,k}(n) \tag{7.3-7}$$

通常，我们令 $j_0=0$，并选择 M 为 2 的幂(即 $M=2^J$)，在式(7.3-5)到式(7.3-7)中，求和在 $n=0,1,2,\cdots,M-1$，$j=0$，$1,2,\cdots,J-1$ 和 $k=0,1,2,\cdots,2^j-1$ 上执行。对于哈尔小波，变换中采用的离散尺度和小波函数(即基函数)与 7.1.3 节中的 $M\times M$ 哈尔变换矩阵的行相对应。变换本身由 M 个系数组成，最小尺度是 0，最大尺度是 $J-1$。出于对 7.3.1 节的注释和例 7.6 中说明的原因，式(7.3-5)和式(7.3-6)中定义的系数通常分别称为近似系数和细节系数。

510 ~ 511

式(7.3-5)到式(7.3-7)中的 $W_\varphi(j_0, k)$和 $W_\psi(j, k)$，分别对应于前一节中的小波级数展开 $c_{j0}(k)$和 $d_j(k)$(这些变量的改变并不是必需的，但为下一节的连续小波变换中的标准表示提供了铺垫)。注意，级数展开中的积分已用求和代替，而曾在 4.4.1 节的 DFT 中出现的归一化因子$1/\sqrt{M}$ 已加到了正向和反向展开表达式中。该因子也可以在正向展开和反向展开表达式中以 $1/M$ 的形式出现。最后，应记住式(7.3-5)到式(7.3-7)只对正交基和紧框架有效。对于双正交基，式(7.3-5)和式(7.3-6)中的 φ 和 ψ 项必须分别由它们的对偶函数 $\tilde{\varphi}$ 和 $\tilde{\psi}$ 来代替。

例 7.8　计算一维离散小波变换。

为说明式(7.3-5)到式(7.3-7)的使用，考虑 4 点离散函数：$f(0)=1, f(1)=4$，$f(2)=-3$ 和 $f(3)=0$。因为 $M=4$，$J=2$，且由于 $j_0=0$，和 $j=0$ 时 $k=0$，$j=1$ 时 $k=0,1$ 在 $x=0,1,2,3$，$j=0,1$ 上执行求和。我们将使用哈尔尺度函数和小波函数，并假定 $f(x)$的 4 个样本分布在基函数的支撑区上，基函数的宽度为 1。将 4 个样本代入式(7.3-5)，我们得到

$$W_\varphi(0, 0)=\frac{1}{2}\sum_{n=0}^{3} f(n)\varphi_{0,0}(n)=\frac{1}{2}\left[1\cdot 1+4\cdot 1-3\cdot 1+0\cdot 1\right]=1$$

因为对于 $n=0, 1, 2, 3$ 有 $\varphi_{0,0}(n)=1$。注意，这里采用的是哈尔尺度函数对于 $j=0$ 和 $k=0$ 的均匀间隔采样。这些值对应于 7.1.3 节的哈尔变换矩阵 $\boldsymbol{H}_4$ 的第一行。继续使用式(7.3-6)和相似间隔的样本 $\psi_{j,k}(x)$，它对应于 $\boldsymbol{H}_4$ 的第 2, 3, 4 行，我们得到

$$W_\psi(0, 0)=\frac{1}{2}\left[1\cdot 1+4\cdot 1-3\cdot(-1)+0\cdot(-1)\right]=4$$

$$W_\psi(1, 0)=\frac{1}{2}\left[1\cdot\sqrt{2}+4\cdot(-\sqrt{2})-3\cdot 0+0\cdot 0\right]=-1.5\sqrt{2}$$

$$W_\psi(1, 1)=\frac{1}{2}\left[1\cdot 0+4\cdot 0-3\cdot\sqrt{2}+0\cdot(-\sqrt{2})\right]=-1.5\sqrt{2}$$

因此，这个简单 4 点样本函数的离散小波变换与哈尔小波及尺度函数的关系就是$\{1, 4, -1.5\sqrt{2}, -1.5\sqrt{2}\}$，这里，变换系数是按计算顺序排列的。

式(7.3-7)可使我们由其变换来重建原始函数。重复求和，我们得到

$$f(n)=\frac{1}{2}\left[W_\varphi(0, 0)\varphi_{0,0}(n)+W_\psi(0, 0)\psi_{0,0}(n)+W_\psi(1, 0)\psi_{1,0}(n)+W_\psi(1, 1)\psi_{1,1}(n)\right]$$

$n=0, 1, 2, 3$。例如，如果 $n=0$，则有

$$f(0)=\frac{1}{2}\left[1\cdot 1+4\cdot 1-1.5\sqrt{2}\cdot(\sqrt{2})-1.5\sqrt{2}\cdot 0\right]=1$$

与正变换情况一样，尺度函数和小波函数的均匀间隔采样也用于反变换的计算。

上例中的 4 点 DWT 是 $f(n)$的一个二尺度分解，即 $j=\{0, 1\}$。基本假设是开始尺度 j_0 为 0，但也可以使用其他的开始尺度。作为练习(见习题 7.16)，请读者在开始尺度为 1 时，计算单尺度变换$\{2.5\sqrt{2}, -1.5\sqrt{2}, -1.5\sqrt{2}, -1.5\sqrt{2}\}$的结果。这样，式(7.3-5)和式(7.3-6)就定义了“一族”开始尺度 j_0 不同的变换。

7.3.3　连续小波变换

离散小波变换的自然延伸是连续小波变换(CWT)，连续小波变换将一个连续函数变换为两个连续变量(平移和尺度)的高冗余度函数。变换结果在时间-频率分析时很容易解释并有很大的价值。虽然我们的兴趣在离散图像上，但出于完整性的考虑，这里也覆盖了连续变换。

511~513

连续平方可积函数$f(x)$的连续小波变换与实数值小波$\psi(x)$的关系定义为

$$W_\psi(s,\tau)=\int_{-\infty}^{\infty} f(x)\psi_{s,\tau}(x)\mathrm{d}x \tag{7.3-8}$$

式中，

$$\psi_{s,\tau}(x)=\frac{1}{\sqrt{s}}\psi\left(\frac{x-\tau}{s}\right) \tag{7.3-9}$$

s和τ分别称为尺度参数和平移参数。给定$W_\psi(s,\tau)$，可以使用连续小波反变换求得$f(x)$：

$$f(x)=\frac{1}{C_\psi}\int_0^{\infty}\int_{-\infty}^{\infty} W_\psi(s,\tau)\frac{\psi_{s,\tau}(x)}{s^2}\mathrm{d}\tau\,\mathrm{d}s \tag{7.3-10}$$

式中，

$$C_\psi=\int_{-\infty}^{\infty}\frac{|\Psi(\mu)|^2}{|\mu|}\mathrm{d}\mu \tag{7.3-11}$$

$\Psi(\mu)$是$\psi(x)$的傅里叶变换。只要满足所谓的允许条件$C_\psi<\infty$ (Grossman and Morlet[1984])，式(7.3-8)到式(7.3-11)就定义了反变换。大多数情况下，这仅意味着$\Psi(0)=0$，且当使得$C_\psi<\infty$，$u\to\infty$时以足够快的速度使$\Psi(\mu)\to 0$。

上面这些等式使我们想起了它们的离散形式，即式(7.2-19)、式(7.3-1)、式(7.3-3)、式(7.3-6)和式(7.3-7)。注意下列相似性：

1. 连续变换参数τ取代了整数平移参数k。
2. 连续尺度参数s与二进制尺度参数2^j相反。这是由于s出现在式(7.3-9)中$\psi((x-\tau)/s)$的分母上。这样，连续变换中使用的小波在$0<s<1$时就被压缩或宽度减小，在$s>1$时就被扩大或展开。小波尺度和传统意义上的频率表示关系是相反的。
3. 连续变换类似于级数展开［见式(7.3-1)］或开始尺度$j_0=-\infty$的离散变换［见式(7.3-6)］。根据式(7.2-26)，这消除了明显的尺度函数间的联系，因此函数仅用小波项来表示。
4. 与离散变换类似，连续变换可视为一组变换系数$\{W\psi(s,\tau)\}$，它度量$f(x)$与一组基函数$\{\psi_{s,\tau}(x)\}$的相似性。但在连续情况下，两个集合都是无穷的。由于$\psi_{s,\tau}(x)$具有实数值，且$\psi_{s,\tau}(x)=\psi^*_{s,\tau}(x)$，所以式(7.3-8)中的每个系数都是$f(x)$和$\psi_{s,\tau}(x)$的内积$\langle f(x),\psi_{s,\tau}(x)\rangle$。

例7.9　一维连续小波变换。
墨西哥草帽小波

$$\psi(x)=\left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)(1-x^2)\mathrm{e}^{-x^2/2} \tag{7.3-12}$$

由其独特的波形而得名［见图7.16(a)］。它与高斯概率函数的二阶导数成正比，均值为零，并且是紧支撑的(即当$|x|\to\infty$时迅速衰减)。虽然它满足连续可逆变换存在的容许性要求，但没有相关的尺度函数，且计算的变换得不到正交分析。其最显著的特性是对称性和式(7.3-12)显式表达的存在性。

图7.16(a)中的连续一维函数是两个墨西哥草帽小波的和：

$$f(x)=\psi_{1,10}(x)+\psi_{6,80}(x)$$

其傅里叶频谱如图7.16(b)所示，该谱揭示了尺度化小波和傅里叶频段之间的紧密联系。频谱中的两个宽频段(或峰值)对应于该函数的两个类高斯扰动。

513~514

图7.16(c)显示了图7.16(a)中关于墨西哥草帽小波函数的CWT的一部分($1\leqslant s\leqslant 10$且$\tau\leqslant 100$)。与

图 7.16(b) 中的傅里叶频谱不同，它同时给出了空间域和频率域信息。注意，例如，当 $s=1$ 时，变换在 $\tau=10$ 时达到最大值，这对应于 $f(x)$ 的分量 $\psi_{1,10}(x)$ 的位置。因为变换为 $f(x)$ 及其计算出的小波之间的相似性提供了客观评价，因此很容易理解它是如何用于特征检测的。我们仅需要与感兴趣的特征相匹配的小波。根据图7.16(d) 中的灰度图也可以得出类似的结论，其中变换的绝对值 $|W\psi(s,\tau)|$ 用黑白之间的灰度来显示。注意，连续小波变换将一个一维函数变成一个二维结果。

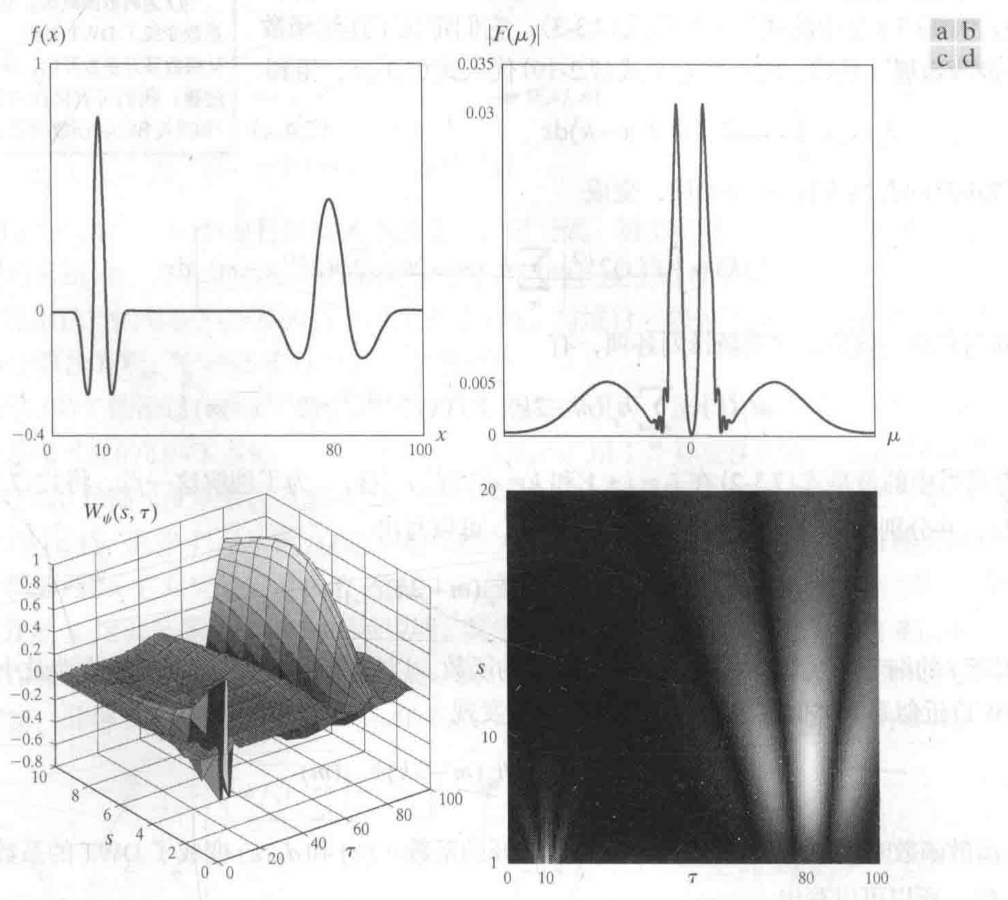


图 7.16　连续小波变换［(c) 和 (d)］和一个连续一维函数 (a) 的傅里叶谱 (b)

7.4　快速小波变换

快速小波变换(FWT) 是一种实现离散小波变换(DWT) 的高效计算，该变换使用了相邻尺度 DWT 系数间的一种幸运且令人惊讶的关系。它也称为 Mallat 人字形算法 (Mallat[1989a, 1989b])，FWT 类似于 7.1.2 节中的 2 子带的子带编码方案。

式 (7.4-1) 是 7.2.2 节中的式 (7.2-18)。

再次考虑多分辨率的详细等式

$$\varphi(x)=\sum_n h_\varphi(n)\sqrt{2}\varphi(2x-n) \tag{7.4-1}$$

用 2^j 对 x 尺度化，用 k 对它平移，并令 $m=2k+n$，给出

$$\begin{aligned}\varphi(2^j x-k)&=\sum_n h_\varphi(n)\sqrt{2}\varphi\big(2(2^j x-k)-n\big)=\sum_m h_\varphi(n)\sqrt{2}\varphi(2^{j+1}x-2k-n)\\&=\sum_m h_\varphi(m-2k)\sqrt{2}\varphi(2^{j+1}x-m)\end{aligned} \tag{7.4-2}$$

514 ~ 515

注意，尺度向量 h_φ 视为用来将 $\varphi(2^j x-k)$展开为尺度 $j+1$ 的尺度函数之和的“加权”。从式(7.2-28)开始的一系列类似操作为 $\psi(2^j x-k)$提供了一个相似的结论，即

$$\psi(2^j x - k) = \sum_m h_\psi(m-2k)\sqrt{2}\varphi(2^{j+1}x - m) \tag{7.4-3}$$

其中，式(7.4-2)中的尺度向量 $h_\varphi(n)$对应于式(7.4-3)中的小波向量 $h_\psi(n)$。

现在考虑7.3.1节中的式(7.3-2)和式(7.3-3)。它们定义了连续函数 $f(x)$的小波级数展开系数。将小波定义式(7.2-19)代入式(7.3-3)，得到

> 当 f 为离散函数时，小波级数展开系数变成了 DWT 系数。这里，我们从级数展开系数开始，以便简化推导过程；我们可直接代入早期的结果(如尺度和小波函数的定义)。

$$d_j(k) = \int f(x)2^{j/2}\psi(2^j x - k)\mathrm{d}x \tag{7.4-4}$$

使用式(7.4-3)的右端代替 $\psi(2^j x-k)$后，变成

$$d_j(k) = \int f(x)2^{j/2}\left[\sum_m h_\psi(m-2k)\sqrt{2}\varphi(2^{j+1}x - m)\right]\mathrm{d}x \tag{7.4-5}$$

交换求和与积分的顺序，并重新排列各项，有

$$d_j(k) = \sum_m h_\psi(m-2k)\left[\int f(x)2^{(j+1)/2}\varphi(2^{j+1}x - m)\right] \tag{7.4-6}$$

其中，方括号中的量是式(7.3-2)在 $j_0=j+1$ 和 $k=m$ 时的 $c_{j0}(k)$。为了理解这一点，将式(7.2-10)代入式(7.3-2)，并分别用 $j+1$ 和 m 代替 j_0 和 k。因此，可以写出

$$d_j(k) = \sum_m h_\psi(m-2k)c_{j+1}(m) \tag{7.4-7}$$

注意，尺度 j 的细节系数是尺度 $j+1$ 的近似系数的函数。以式(7.4-2)和式(7.3-2)作为涉及小波级数展开(和 DWT)近似系数类似推导的起点，我们同样发现

$$c_j(k) = \sum_m h_\varphi(m-2k)c_{j+1}(m) \tag{7.4-8}$$

当 $f(x)$为离散函数时(见7.3.2节)，因为小波级数展开的系数 $c_j(k)$ 和 $d_j(k)$ 变成了 DWT 的系数 $W_\varphi(j,k)$ 和 $W_\psi(j,k)$，所以可以写出

$$W_\psi(j,\ k) = \sum_m h_\psi(m-2k)W_\psi(j+1,\ m) \tag{7.4-9}$$

$$W_\varphi(j,\ k) = \sum_m h_\varphi(m-2k)W_\varphi(j+1,\ m) \tag{7.4-10}$$

式(7.4-9)和式(7.4-10)揭示了相邻尺度的 DWT 系数间的明显关系。将这些结果与式(7.1-7)进行比较，我们看到 $W_\varphi(j,k)$ 和 $W_\psi(j,k)$，尺度 j 的近似和细节系数可用 $W_\phi(j+1,k)$ 和尺度 $j+1$ 的近似系数，分别与顺序倒置尺度和小波向量 $h_\varphi(-n)$ 与 $h_\psi(-n)$ 进行卷积操作，然后对结果下取样来计算。图 7.17 以方框图的形式总结了这些操作。注意，这个方框图与图 7.6 中的 2 带度子带编码和解码系统在 $h_0(n)=h_\varphi(-n)$ 和 $h_1(n)=h_\psi(-n)$ 时的分析部分相同。因此，可以写出

$$W_\psi(j,\ k) = h_\psi(-n)\star W_\varphi(j+1,\ n)\Big|_{n=2k,\ k\geq 0} \tag{7.4-11}$$

和

$$W_\varphi(j,\ k) = h_\varphi(-n)\star W_\varphi(j+1,\ n)\Big|_{n=2k,\ k\geq 0} \tag{7.4-12}$$

516 ~ 517

其中，卷积在 $n=2k, k\geqslant 0$ 时直接求值。如例 7.10 中所述，非负偶数项求卷积等价于滤波和基 2 下取样。

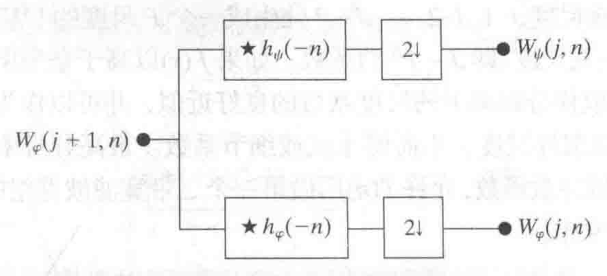


图 7.17　一个 FWT 分析滤波器组

> 如果将式(7.4-9)中的 $h_\varphi(m-2k)$ 重写为 $h_\varphi(-(2k-m))$，我们会发现第一个减号负责颠倒顺序［见式(7.1-6)］，$2k$ 负责子取样［见式(7.1-2)］，m 是卷积的哑变量［见式(7.1-7)］。

式(7.4-11)和式(7.4-12)是计算快速小波变换的定义式。对于长度为 $M=2^J$ 的序列，所涉及的数学运算次数约为$O(M)$，即乘法和加法的次数与输入序列的长度呈线性关系，因为在图7.17 中，用 FWT 分析滤波器组执行的卷积所涉及的乘法和加法的次数，与进行卷积序列的长度成正比。这样，用 FWT 就比用 FFT 算法有利，它约要求 $O(M\log_2 M)$次操作。

为总结 FWT 的论述，我们注意到图 7.17 中的滤波器组可“迭代”地产生多级结构，以便计算两个或多个连续尺度的 DWT 系数。例如，图7.18(a)显示了用于计算变换的两个最高尺度系数的二级滤波器组。注意，最高尺度系数假定为函数本身的取样值，即 $W_\varphi(J,n)=f(n)$，其中 J 表示最高尺度［根据 7.2.2 节，$f(x)\in V_J$，其中 V_J是函数$f(x)$所在的尺度空间］。图 7.18(a)中的第一个滤波器组将原始函数分解为一个低通函数［它对应于尺度系数 $W_\varphi(J-1,n)$的近似分量］和一个高通函数［它对应于系数 $W_\psi(J-1,n)$的细节分量］。这就是图 7.18(b)的图示说明，其中尺度空间 V_J 被分成小波子空间 W_{J-1} 和尺度子空间 V_{J-1}。原始函数的频谱被分成两个半带宽分量。图7.18(a)中的第二个滤波器组将频谱和子空间 V_{J-1}，较低的半带宽，分解成 1/4 带宽子空间 W_{J-2} 和 V_{J-2}，分别对应于 DWT 系数 $W_\psi(J-2,n)$和 $W_\varphi(J-2,n)$。

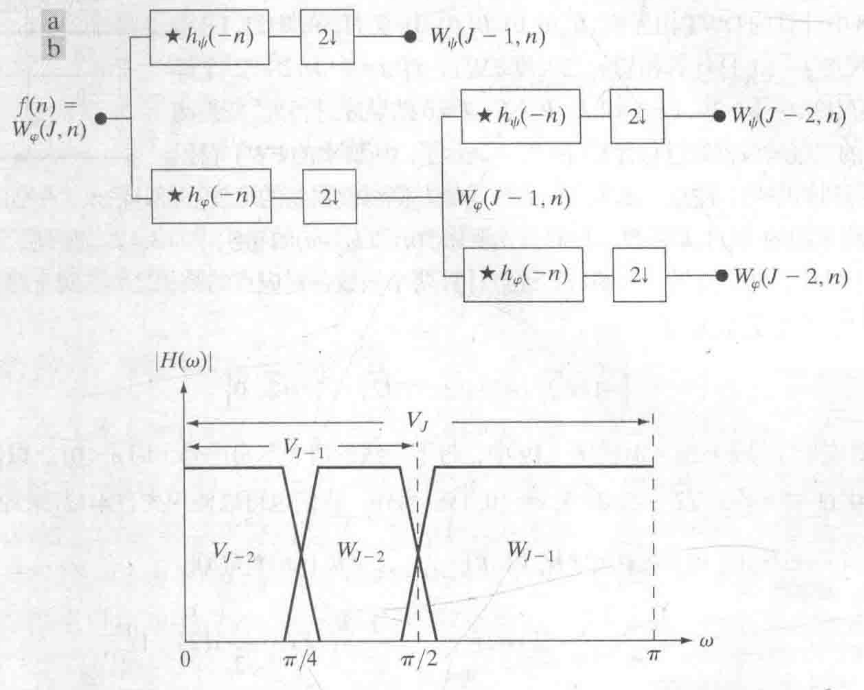


图 7.18　(a)一个二级或二尺度 FWT 分析滤波器组；(b)其频率分离特性

517 ~ 518

图7.18(a)中的二级滤波器组很容易扩展到任意尺度。例如，第三个滤波器组将在 $W_\varphi(J-2,n)$ 系数上操作，将尺度空间 V_{J-2} 分成两个 1/8 带宽的子空间 W_{J-3} 和 V_{J-3}。通常，我们选择 $f(x)$的 2^J 个样本，并用 P 个滤波器组(如图 7.17 所示)在尺度 $J-1, J-2, \cdots, J-P$ 处生成一个 P 尺度的 FWT。首先计算最高尺度(即 $J-1$)的系数，最后计算最低尺度(即 $J-P$)的系数。如果 $f(x)$以高于奈奎斯特率的取样率取样（通常如此），那么其样本是该取样分辨率下的尺度系数的良好近似，并可以作为起始的高分辨率尺度系数的输入。换句话说，在该取样尺度，不需要小波或细节系数。最高分辨率的尺度函数作为式(7.3-5)和式(7.3-6)中的单位离散冲激函数，允许 $f(n)$用做第一个二带宽滤波器组的尺度(近似)输入(Odegard, Gopinath and Burrus[1992])。

例 7.10 计算一维快速小波变换。

为解释前面的概念，考虑例 7.8 中的离散函数 $f(n) = \{1, 4, -3, 0\}$。如例 7.8 那样，我们将计算基于哈尔尺度和小波函数的变换。但这里不再像例 7.8 中的 DWT 那样直接使用基函数，而是使用例7.5 和例 7.6 中的相应尺度和小波向量：

$$h_\varphi(n) = \begin{cases} 1/\sqrt{2}, & n = 0.1 \\ 0, & \text{其他} \end{cases} \tag{7.4-13}$$

和

$$h_\psi(n) = \begin{cases} 1/\sqrt{2}, & n = 0 \\ -1/\sqrt{2}, & n = 1 \\ 0, & \text{其他} \end{cases} \tag{7.4-14}$$

这些是用于建立 FWT 滤波器组的函数；它们提供了滤波器系数。注意，因为哈尔尺度函数和小波函数是归一化正交的，式(7.1-14)可用在类似表7.2 中的 $h_\varphi(n)$ 单个原型滤波器，产生 FWT 滤波器系数，它对应于式(7.1-14)中的 $g_0(n)$。

表7.2 $h_\varphi(n)$ 的正交归一化哈尔滤波器系数

n	$h_\varphi(n)$
0	$1/\sqrt{2}$
1	$1/\sqrt{2}$

因为例 7.8 中计算的DWT 由元素$\{W_\varphi(0,0), W_\psi(0,0), W_\psi(1,0), W_\psi(1,1)\}$组成，我们将对尺度 $j=\{0,1\}$计算相应的二尺度 FWT，即 $J=2$(有 $2^J=2^2$个样本)且 $P=2$(按尺度 $J-1=2-1=1$ 和 $J-P=2-2=0$ 的顺序进行)。变换使用图 7.18(a)中的二级滤波器组进行计算。图 7.19 显示了由所要求的 FWT 卷积和下取样操作得到的序列。注意，函数 $f(n)$自身是最左侧滤波器组的尺度(近似)输入。例如，为计算出现在图 7.19 中上支路末端的 $W_\psi(1,k)$系数，我们首先要做 $f(n)$和 $h_\psi(-n)$的卷积。如 3.4.2 节所述，这要求关于原点翻转其中的一个函数，并滑过另一个函数，逐点计算两个函数各对应点的乘积之和。对于序列$\{1, 4, -3, 0\}$和$\{-1/\sqrt{2}, 1/\sqrt{2}\}$，产生的结果为

$$\left\{-1/\sqrt{2},\ -3/\sqrt{2},\ 7/\sqrt{2},\ -3/\sqrt{2},\ 0\right\}$$

其中，第二项对应于序号 $k=2n=0$(在图 7.19 中，带下画线的值表示负序号，即 $n<0$)。以偶序号点下取样时，我们得到 $W_\psi(1,k)=\{-3/\sqrt{2}, -3/\sqrt{2}\}, k=\{0,1\}$。另外，我们也可以使用式(7.4-12)来计算：

$$W_\psi(1,k) = h_\psi(-n) \star W_\varphi(2,n)\big|_{n=2k,\ k\geq 0} = h_\psi(-n) \star f(n)\big|_{n=2k,\ k\geq 0}$$

$$= \sum_l h_\psi(l-2k)x(l)\bigg|_{k=0,1} = \frac{1}{\sqrt{2}}x(2k) - \frac{1}{\sqrt{2}}x(2k+1)\bigg|_{k=0,1}$$

518 ~ 520

这里，我们用 $2k$ 代替卷积中的 n，并采用 l 作为卷积的虚变量(即置换相关的两个序列)。展开的和中只有

两项，因为反转的小波向量 $h_\psi(-n)$中只有两个非零值。替代 $k=0$，我们得出 $W_\psi(1,0)=-3/\sqrt{2}$；对于 $k=1$，我们得到 $W_\psi(1,1)=-3/\sqrt{2}$。这样，滤波并下取样后的序列为 $\{-3/\sqrt{2},-3/\sqrt{2}\}$，它与先前的结果一致。剩下的卷积和下取样操作可按类似的方法进行。

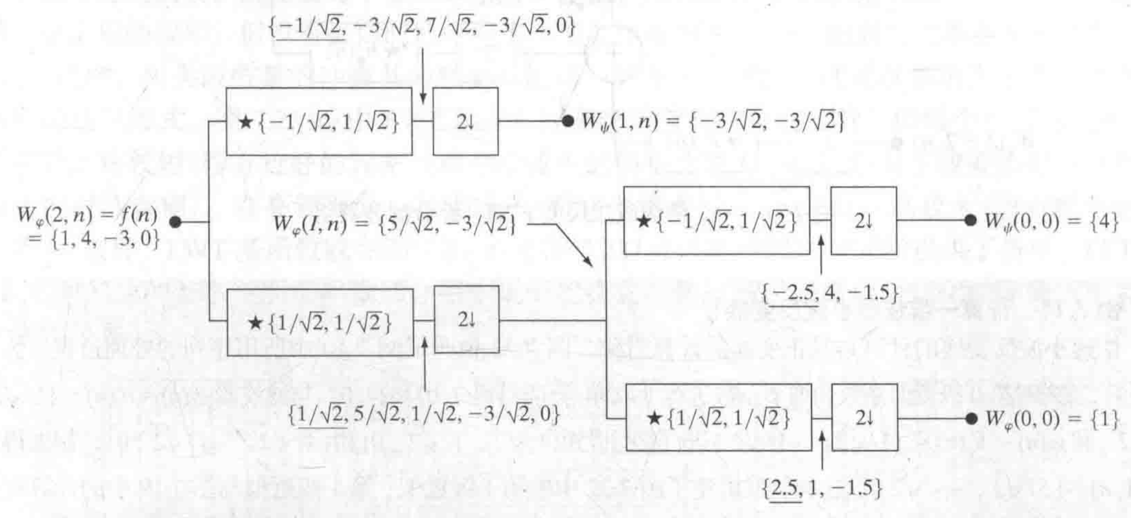


图 7.19　使用哈尔尺度和小波向量计算序列{1, 4, −3, 0}的一个二尺度快速小波变换

如可能期望的那样，从正变换的结果重建 $f(n)$的快速反变换也可以表达出来。这称为快速小波反变换(FWT^{-1})，它使用正变换中采用的尺度和小波向量，以及第 j 级的近似和细节系数，来生成第 $j+1$ 级的近似系数。注意图 7.17 中的 FWT 分析滤波器组和图 7.6(a)中的 2 带宽子带分析部分之间的相似性，我们可以立即得出所需的 FWT^{-1} 综合滤波器组。图 7.20 详细描述了它的结构，该结构与图 7.6(a)中的 2 带宽子带编码和解码系统的综合部分相同。7.1.2 节中的式(7.1-14)定义了相关的综合滤波器。如该节解释的那样，完美重建(对于二子带正交滤波器)要求对于 $i=\{0,1\}$有 $g_i(n)=h_i(-n)$。也就是说，综合滤波器和分析滤波器彼此之间必须是顺序相反的形式。因为 FWT 分析滤波器(见图7.17)是 $h_0(n)=h_\varphi(-n)$和 $h_1(n)=h_\psi(-n)$，所以要求的 FWT^{-1} 综合滤波器为 $g_0(n)=h_0(-n)=h_\varphi(n)$和 $g_1(n)=h_1(-n)=h_\psi(n)$。但应记住，也可以使用双正交分析滤波器和综合滤波器，但它们彼此之间形式上顺序不相反。双正交分析滤波器和综合滤波器是根据式(7.1-10)和式(7.1-11)交叉调制的。

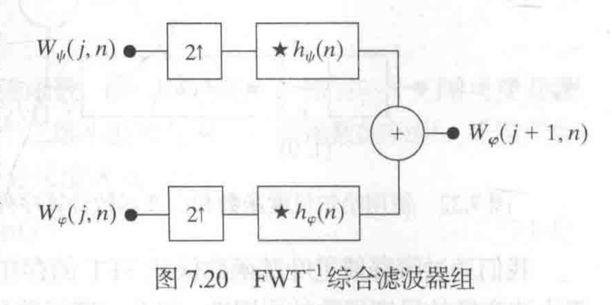


图 7.20　FWT^{-1} 综合滤波器组

图 7.20 中的 FWT^{-1} 滤波器组执行下述计算：

$$W_\varphi(j+1,k)=h_\varphi(k)\star W_\varphi^{2\uparrow}(j,k)+h_\psi(k)\star W_\psi^{2\uparrow}(j,k)\Big|_{k\geq 0} \tag{7.4-15}$$

式中，$W^{2\uparrow}$ 代表基 2 上取样［即在式(7.1-1)定义的 W 中插入 0，使其长度变为原来的 2 倍］。上取样的系数与 $h_\varphi(n)$和 $h_\psi(n)$进行卷积完成滤波，并相加产生较高尺度的近似。本质上，这创建了 $f(n)$的较好近似，含有较多的细节和较高的分辨率。与 FWT 正变换类似，反变换滤波器组可以如图 7.21 所示的那样进行迭代。其中，为了计算 FWT^{-1} 重建的最后两个尺度，描绘了一个二尺度结构。这种系数合并过程可以扩展到任意数量的尺度，并保证序列 $f(n)$的完美重建。

> 记住，像在金字塔编码中一样(见 7.1.1 节)，小波变换可以按用户指定数量的尺度计算。例如，对于一幅大小为 $2^J\times 2^J$ 的图像，有 $1+\log_2 J$ 个可能的尺度。

520
~
521

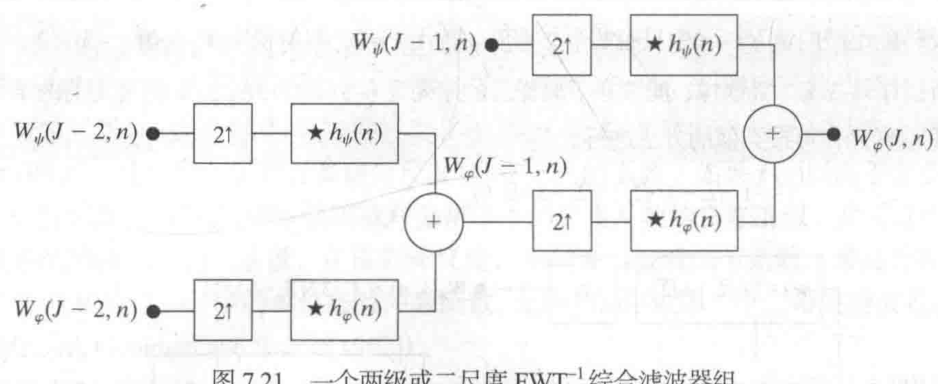


图 7.21 一个两级或二尺度 FWT^{-1} 综合滤波器组

例 7.11 计算一维快速小波反变换。

快速小波反变换的计算与其正变换的计算对称。图 7.22 说明了例 7.10 中所用序列的处理过程。为开始计算，首先对第 0 级近似系数和细节系数进行上取样，分别得到$\{1, 0\}$和$\{4, 0\}$。与滤波器 $g_0(n)=h_\varphi(n)=\{1/\sqrt{2}, 1/\sqrt{2}\}$和 $g_1(n)=h_\psi(n)=\{1/\sqrt{2}, -1/\sqrt{2}\}$进行卷积得到$\{1/\sqrt{2}, 1/\sqrt{2}, 0\}$和$\{4/\sqrt{2}, -4/\sqrt{2}, 0\}$，相加得到 $W_\varphi(1,n)=\{5/\sqrt{2}, -3/\sqrt{2}\}$。这样，就重建了图 7.22 中的第 1 级近似，第 1 级近似与图 7.19 中的计算近似匹配。以这种方式继续，在第二个综合滤波器组的右端形成了 $f(n)$。

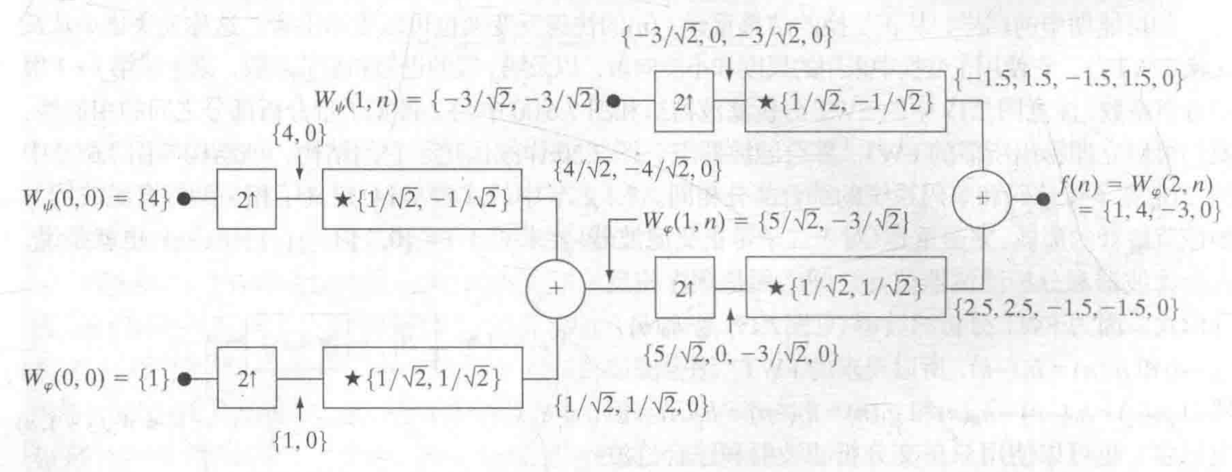


图 7.22 使用哈尔尺度函数和小波函数计算序列$\{1, 4, -1.5\sqrt{2}, -1.52\sqrt{2}\}$的二尺度快速小波反变换

我们通过解释傅里叶基函数保证 FFT 的存在性来结束快速小波变换的讨论，FWT 的存在性取决于小波变换的尺度函数的可用性，以及尺度函数和对应小波的正交性(或双正交性)。这样，式(7.3-12)中墨西哥草帽小波的确没有相伴的尺度函数，所以不能用于 FWT 的计算。换句话说，不能为墨西哥草帽小波建立一个如图 7.17 所示的滤波器组，因为它不满足 FWT 方法的基本假设。

最后，我们注意到，在表示函数时，时间和频率通常被视为不同的域，但它们之间存在着不可分割的关系。如果试图同时在时间域和频率域内对函数进行分析，就会遇到如下问题：如果想得到关于时间域的精确信息，就必须接受频率域的某些模糊，反之亦然。这是海森伯(Heisenberg)测不准原理在信息处理中的应用。为了用图示方式说明该原理，函数表示中用到的每个基函数都可粗略地视为时间-频率平面中的一片。该片也称为海森伯单元或海森伯盒，它显示了其所表示的基函数的频率内容，以及该基函数在时域中所处的位置。归一化正交基函数由不重叠的片来表征。

522 ~ 523

图 7.23 分别显示了时间-频率片：(a)冲激函数(即传统的时间域)基；(b)正弦函数(FFT)基；

(c) FWT 基。图7.23 (a) 到 (c) 的每一片都是一个方形区域；区域的高和宽定义了可用基函数表示的函数的频率和时间特性。注意，图 7.23 (a) 中的标准时域基指明了事件发生的时刻，但不提供频率信息 [图7.23 (a) 中每个矩形的宽度应被考虑为时间上的一个瞬间)。这样，为了用冲激基函数表示一个单一频率的正弦波，就需要每个基函数。另一方面，图7.23 (b) 中的正弦基指出了发生较长时间的事件中出现的频率，但没有提供时间分辨率 [图7.23 (b) 中每个矩形的高度应考虑为一个单一的频率)。这样，由无限数量的冲激基函数表示的单一频率的正弦波，就可以表示为包含一个正弦基函数的展开形式。图 7.23 (c) 中 FWT 片的时间和频率分辨率是变化的，但每个片的面积相同。在低频处，片较短 (即有较好的频率分辨率或较少的频率含糊)，但较宽 (对于较差的时间分辨率或较大的时间含糊)。在高频部分，片宽较小 (时间分辨率提高)，并且片高较大 (表示频率分辨率下降)。这样，FWT 基函数就在图 7.23 (a) 和图 7.23 (b) 这两种限制情况间提供了折中。FFT 和 FWT 之间的这种基本差别在本章的介绍中说明过，它在频率随时间变化而变化的非稳态函数的分析中很重要。

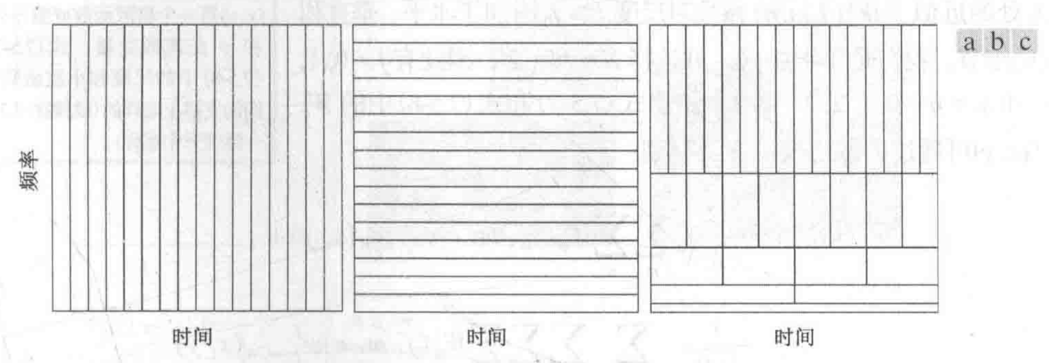


图 7.23　与 (a) 取样数据，(b) FFT 和 (c) FWT 相关的基函数的时间-频率片。注意，图 (c) 中等高度矩形的水平条带表示 FWT 的尺度

7.5 二维小波变换

前述一维变换可很容易地推广到像图像这样的二维函数。在二维情况下，需要一个二维尺度函数 $\varphi(x, y)$ 和三个二维小波 $\psi^H(x, y)$, $\psi^V(x, y)$ 和 $\psi^D(x, y)$。每个二维小波都是两个一维函数的乘积。排除产生一维结果的乘积，如 $\varphi(x)\psi(x)$，4 个剩下的乘积产生可分尺度函数

$$\varphi(x, y) = \varphi(x)\varphi(y) \tag{7.5-1}$$

和可分的"方向敏感"小波

$$\psi^H(x, y) = \psi(x)\varphi(y) \tag{7.5-2}$$

$$\psi^V(x, y) = \varphi(x)\psi(y) \tag{7.5-3}$$

$$\psi^D(x, y) = \psi(x)\psi(y) \tag{7.5-4}$$

这些小波度量函数的变化——图像的灰度变化——沿不同方向的变化：ψ^H 度量沿列方向的变化 (例如水平边缘)，ψ^V 响应沿行方向的变化 (例如垂直边缘)，ψ^D 度量对应对角线方向的变化。方向敏感是式 (7.5-2) 到式 (7.5-4) 中可分性的自然结果；它并不会增加本节讨论的二维变换的计算复杂性。

给定可分二维尺度函数和小波函数，一维 DWT 可直接扩展到二维。我们首先定义一个尺度和平移基函数：

523 ~ 524

$$\varphi_{j,m,n}(x,y)=2^{j/2}\varphi(2^j x-m,\ 2^j y-n) \tag{7.5-5}$$

$$\psi^i_{j,m,n}(x,y)=2^{j/2}\psi^i(2^j x-m,\ 2^j y-n),\ i=\{H,\ V,\ D\} \tag{7.5-6}$$

式中，上标 i 指出式(7.5-2)到式(7.5-4)中的方向小波。与指数不同，i 是假定值 H, V 和 D 的上标。于是，大小为 $M\times N$ 的图像 $f(x,y)$的离散小波变换是

$$W_\varphi(j_0,\ m,\ n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,\ y)\varphi_{j_0,m,n}(x,\ y) \tag{7.5-7}$$

$$W^i_\psi(j,\ m,\ n)=\frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,\ y)\psi^i_{j,m,n}(x,\ y),\ \ i=\{H,\ V,\ D\} \tag{7.5-8}$$

如同一维情况，j_0 是一个任意的开始尺度，$W_\varphi(j_0, m, n)$系数定义 $f(x, y)$ 在尺度 J_0 处的近似。$W^i_\psi(j,m,n)$ 系数对尺度 $j\geqslant j_0$ 附加了水平、垂直和对角方向的细节。我们通常令 $j_0=0$，并选择 $N=M=2^J$，因此有 $j=0, 1, 2,\cdots, J-1$ 和 $m=n=0, 1, 2,\cdots, 2^j-1$。给出式(7.5-7)和式(7.5-8)中的 W_φ 和 W^i_ψ，$f(x,y)$可通过离散小波反变换得到：

> 既然我们正在处理二维图像，故 $f(x,y)$是一个离散函数或值序列，而 x 和 y 是离散变量。式(7.5-7)和式(7.5-8)中的尺度和小波函数是在它们的支撑上取样的(就像在 7.3.2 节的一维情形中那样)。

$$f(x,\ y)=\frac{1}{\sqrt{MN}}\sum_m\sum_n W_\varphi(j_0,\ m,\ n)\varphi_{j_0,m,n}(x,\ y)+ \\ \frac{1}{\sqrt{MN}}\sum_{i=H,V,D}\sum_{j=j_0}^{\infty}\sum_m\sum_n W^i_\psi(j,\ m,\ n)\psi^i_{j,m,n}(x,\ y) \tag{7.5-9}$$

类似于一维离散小波变换，二维 DWT 可以使用数字滤波器和下取样器来实现。利用可分的二维尺度函数和小波函数，我们可以先简单地取 $f(x,y)$的行的一维 FWT，然后，取结果列的一维 FWT。图7.24(a)以方框图的形式显示了这一过程。注意，类似于图7.17 的一维对应部分，二维 FWT“滤波”尺度 $j+1$ 的近似系数来构建尺度 j 的近似和细节系数。但在二维情况下，我们得到了 3 组细节系数——水平、垂直和对角线细节系数。

图7.24(a)的单尺度滤波器组可通过“迭代”(将该近似输出连接到另一个滤波器组输入)方式，在尺度 $j=J-1, J-2,\cdots, J-P$ 中产生 P 尺度变换。如一维情形那样，图像 $f(x,y)$被用做 $W_\varphi(J, m, n)$的输入。它的行与 $h_\varphi(-n)$和 $h_\psi(-n)$卷积，并对它的列下取样，得到两个子图像，它们的水平分辨率以 2 为基下降。高通或细节分量描述了图像垂直方向的高频信息，低通或近似分量包含了低频垂直信息。然后，沿列的方向对两幅子图像进行滤波并下取样，得到 4 幅 1/4 大小的输出子图像——W_φ，W^H_ψ，W^V_ψ 和 W^D_ψ。显示于图7.24(b)中间的这些子图像是 $f(x,y)$与式(7.5-1)到式(7.5-4)中的二维尺度函数与小波函数的内积，紧接着在每个维度进行基 2 下取样。滤波处理的两次迭代在图 7.24(b)的最右侧产生二尺度分解。

> W_φ，W^H_ψ，W^V_ψ 和 W^D_ψ 按图 7.24(b)中的方式排列。计算出的每个尺度将替代它们之前基于的尺度近似。

图7.24(c)显示了刚刚描述过的反向处理的综合滤波器组。如所期望的那样，重建算法类似于一维情况。在每次迭代中，四尺度 j 的近似和细节子图像被上取样，并与两个一维滤波器卷积——一个在子图像的列上进行操作，另一个在行上进行操作。将结果相加得到尺度 $j+1$ 的近似，重复该处理直到重建原图像。

524 ~ 526

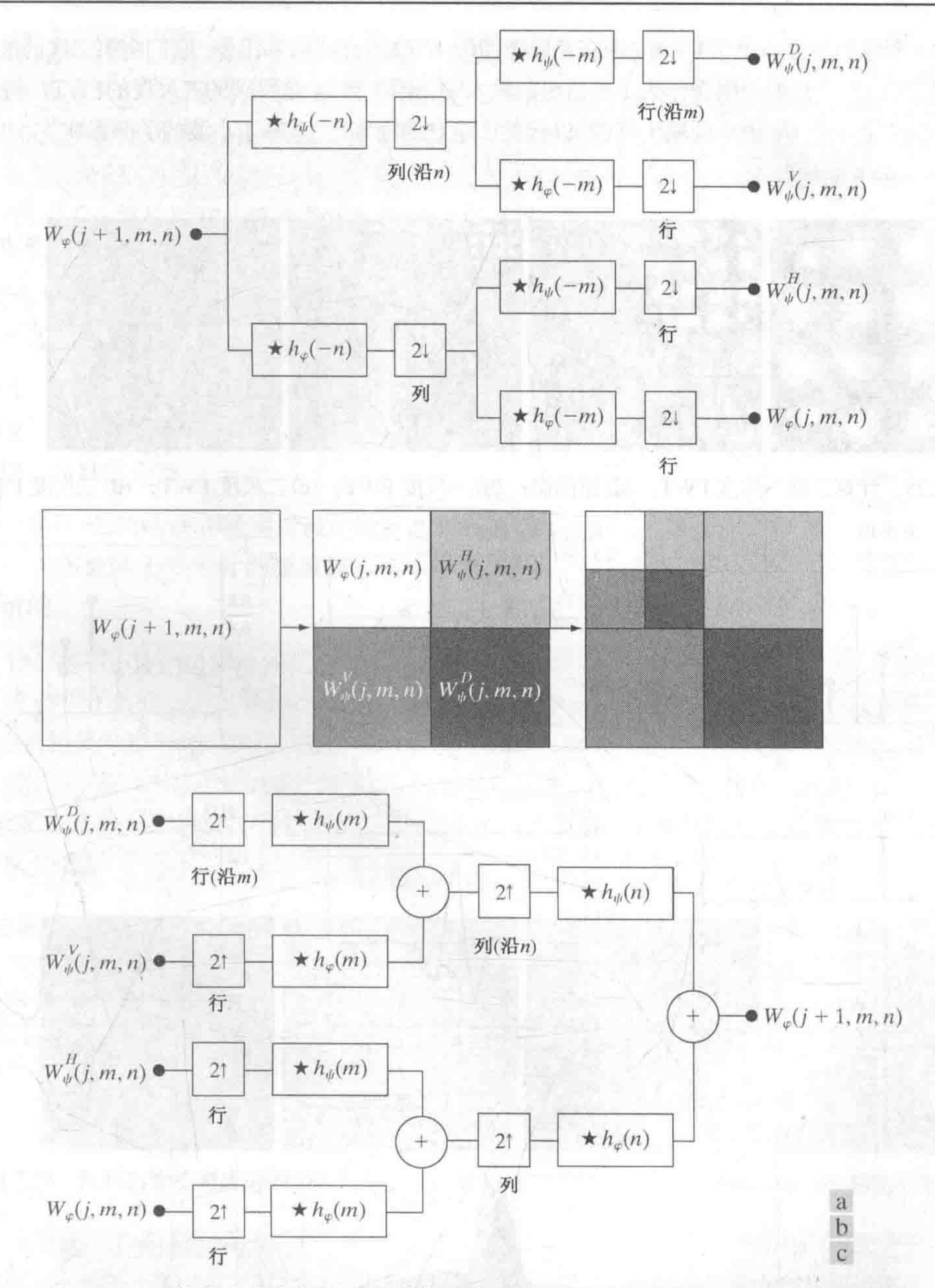


图 7.24　二维快速小波变换：(a) 分析滤波器组；(b) 分解结果；(c) 综合滤波器组

例 7.12　计算二维快速小波变换。

图 7.25 (a) 是由计算机产生的一幅 128 × 128 的图像，它由黑色背景上的二维类似正弦脉冲组成。该例子的主要目的是说明计算图像的二维 FWT 的机理。图 7.25 (b) 到 (d) 显示了图 7.25 (a) 中图像的三个 FWT。图 7.24 (a) 的二维滤波器组及图 7.26 (a) 和 (b) 中显示的分解滤波器用于产生所有的三个结果。

> 本例中使用的尺度和小波向量稍后描述。此处，我们的重点在于变换计算的机理，它与采用的滤波器系数无关。

图 7.25 (b) 显示了图 7.25 (a) 中图像的一尺度 FWT。为了计算该变换，原图像作为图 7.42 (a) 的滤波器组的输入。然后，4 个 1/4 大小的分解输出 (即近似、水平、垂直和对角线) 按照图 7.24 (b) 的排列产生图 7.25 (b) 中的图像。类似的过程可用来产生图 7.25 (c) 中的二尺度 FWT，但滤波器组的输入变为图 7.25 (b) 左上角所示 1/4 大小的近似子图像。如我们在图 7.25 (c) 中看到的那样，

525 ~ 526

1/4 大小的子图像然后由 4 个 1/4 大小(现在是原图像的 1/16)的分解结果代替，它们由第二次滤波产生。最后，当图 7.25(c)左上角的子图像作为滤波器组的输入时，图 7.25(d)是得到的三尺度的FWT。每次通过滤波器组，便会产生 4 个 1/4 大小的输出图像，以代替产生它的输入。注意基于小波的子图像 W_ψ^H，W_ψ^V 和 W_ψ^D 在每个尺度处的方向特性。

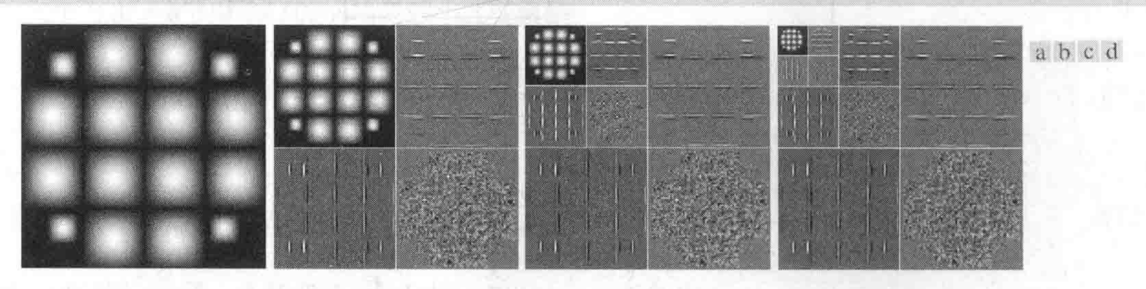


图 7.25　计算二维三尺度 FWT：(a)原图像；(b)一尺度 FWT；(c)二尺度 FWT；(d)三尺度 FWT

图 7.26　四阶对称小波：(a)~(b)分解滤波器；(c)~(d)重建滤波器；(e)一维小波；(f)一维尺度函数；(g)三个二维小波之一，$\psi^V(x, y)$。$h_\varphi(n)$, $0 \leqslant n \leqslant 7$ 的值见表 7.3

527 ~ 528

前面例子中使用的分解滤波器是众所周知的称为对称小波的小波族的一部分。symlet 是对称小波的英文简称。虽然它们不完全对称，但在给定的紧支撑下(Daubechies[1992])被设计为有最小不对

称性和最高消失矩数[①]。图 7.26(e)和图 7.26(f)显示了 4 阶一维对称小波(即小波和尺度函数)。图7.26(a)到图7.26(d)显示了相应的分解和重建滤波器。低通重建滤波器 $g_0(n)=h_\varphi(n)$的系数在 $0 \leqslant n \leqslant 7$ 的情况下由表 7.3 给出。其余归一化正交滤波器的系数可用式(7.1-14)得到。作为小波 $\psi^V(x,y)$的低分辨率图解，图 7.26(g)说明了一维尺度和小波函数是如何结合在一起来形成一个可分二维小波的。

回忆可知，一个函数的紧支撑是指一个区间，该函数在这个区间内有非零值。

表 7.3　$h_\varphi(n)$ 的正交归一化四阶对称小波滤波器系数(Daubechies[1992])

n	$h_\varphi(n)$
0	0.0322
1	−0.0126
2	−0.0992
3	0.2979
4	0.8037
5	0.4976
6	−0.0296
7	−0.0758

我们用两个演示小波在图像处理中的作用的例子来结束本节。如在傅里叶域那样，基本方法是：

步骤 1　计算一幅图像的二维小波变换。

步骤 2　修改变换。

步骤 3　计算反变换。

因为 DWT 的尺度和小波向量被用做低通和高通滤波器，所以大多数基于傅里叶的滤波技术与“小波域”部分是等价的。

例 7.13　基于小波的边缘检测。

图 7.27 提供了前面三个步骤的简单说明。在图7.27(a)中，示于图7.25(c)的离散小波变换的最低尺度近似分量已由将其值置零而消除。如图7.27(b)所示，使用这些改进的系数计算小波反变换时，最终效果是边缘增强，回忆一下在 4.9 节讨论的基于傅里叶的图像锐化结果。注意，尽管它们是相对软的正弦过渡，但信号和背景间的过渡多少还是被很好地描述了。通过将水平细节置零——参见图7.27(c)和图 7.27(d)，我们可以孤立出垂直边缘。

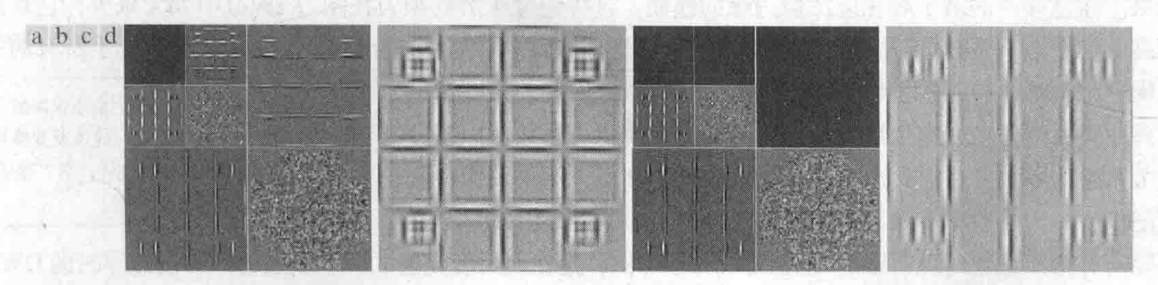


图 7.27　针对边缘检测改进的 DWT：(a)~(c)删除所选系数的二尺度分解；(b)~(d)相应的重建

例 7.14　基于小波的噪声去除。

作为第二个例子，考虑示于图7.28(a)中的人体头部 CT 图像。正如在背景中所看到的那样，图像已被加性白噪声均匀污染。通常，基于小波来对图像去噪(即消除噪声部分)的过程如下：

步骤 1　为了分解，选择一个小波(如哈尔小波、对称小波等)和级别数(尺度)P。然后，计算噪声图像的 FWT。

步骤 2　对细节系数进行阈值处理，即从尺度 $J-1$ 到 $J-P$，选择和应用一个阈值处理细节系数。这可以通过硬阈值来实现，其含义是将绝对值低于阈值的元素置零；或通过软阈值处理实现，即首先将其绝对值比阈值低的元素置零，然后将非零系数标定到接近零。软阈值处理消除了硬阈值处理中(在阈值处)固有的不连续性(见第 10 章关于阈值处理的讨论)。

① 小波 $\psi(x)$的第 k 阶矩是 $m(k)=\int x^k\psi(x)\mathrm{d}x$。0 阶矩影响尺度函数和小波函数的平滑性以及我们将它们表示为多项式的能力。一个 N 阶对称小波有 N 个消失矩。

529 ~ 530

步骤 3　在 $J-P$ 级处使用原始近似系数，并对级 $J-1$ 到 $J-P$ 的改进细节系数计算小波反变换(即执行小波重建)。

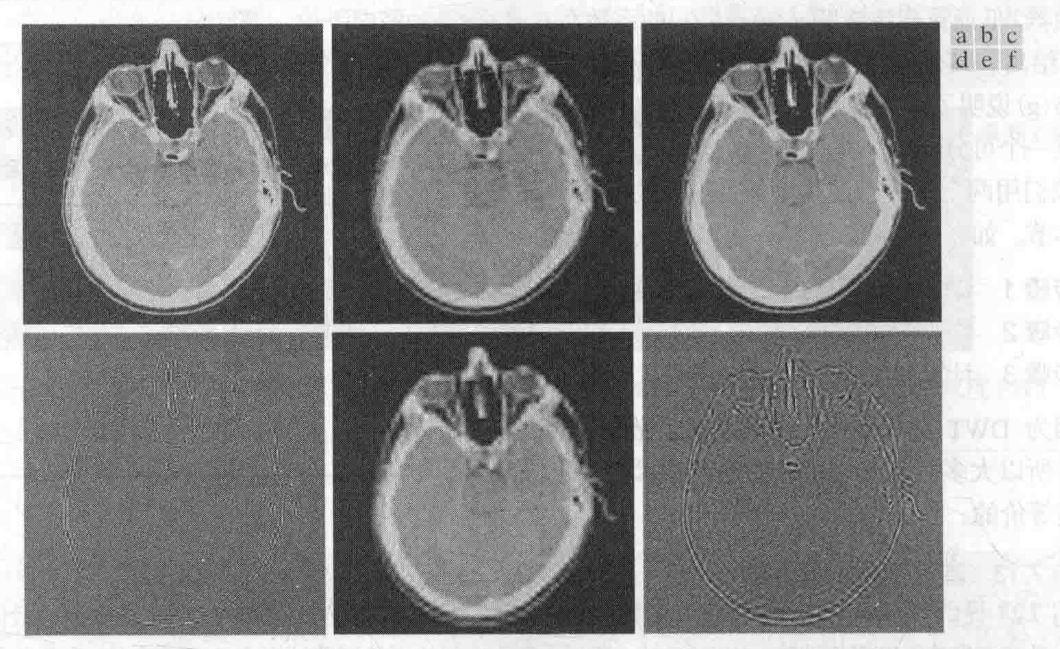


图 7.28　为噪声去除修改 DWT：(a)人的头部的噪声 CT 图像；(b)，(c)和(e)对细节系数进行阈值处理后的各种重建；(d)和(f)在(c)和(e)重建过程中所删除的信息(原图像由 Vanderbilt 大学医学中心提供)

图 7.28(b)显示了使用第四阶对称小波、二尺度($P=2$)及一个交互确定的全局阈值来执行这些操作的结果。注意噪声的减少和相应图像边缘的模糊。这种边缘细节的损失在图 7.28(c)中大大减少，它是由最高分辨率细节系数(不是低分辨率细节的阈值处理)简单置零并重建图像产生的。这里，几乎所有的背景噪声都已被去掉，而边缘仅稍有干扰。图 7.28(d)中的差值图像显示了处理中丢失的信息。这个结果是计算两尺度变换所有系数，但高分辨率细节系数为零的反 FWT 产生的。如所看到的那样，结果图像中包含了原图像中的大多数噪声和某些边缘信息。图 7.28(e)和图 7.28(f)显示了删除所有细节系数的负面影响。也就是说，图 7.28(e)是在两尺度变换的两个级别的细节被置零时的 DWT 重建；图 7.28(f)显示了丢失的信息。注意，图 7.28(f)中的边缘信息明显增加，而图 7.28(e)中的边缘细节相应减少。

在生成图 7.28(d)时，由于仅保留了最高分辨率细节系数，故该反变换是它们对图像的贡献。同样，图 7.28(f)是所有细节系数的贡献。

7.6　小波包

快速小波变换将一个函数分解为宽度呈对数关系的尺度函数和小波函数的和，即(函数的)低频内容使用较窄带宽的(尺度和小波)函数来表示，而高频内容使用较宽带宽的函数来表示。如果沿图 7.23(c)中的时间-频率平面的频率轴来看，这非常明显。每个恒定高度片的水平条带包含单一 FWT 尺度的基函数，向上移动频率轴时，在高度上呈对数增加。如果我们想要更大地控制时间-频率平面的分开度(即在较高频率处频带较小)，则必须对 FWT 加以推广，以便得到一个更为灵活的分解，这种分解称为小波包(Coifman and Wickerhauser[1992])。这个推广的代价是计算复杂性从 FWT 的 $O(M)$ 增大到小波包的 $O(M\log_2 M)$。

530 ~ 532

再次考虑图7.18(a)中的二尺度滤波器组，但此时把分解想象为一个二叉树。图7.29(a)详细叙述了

树的结构，并适当地将 FWT 尺度和小波系数［从图 7.18(a)］连接到它的节点。根节点被赋予了最高尺度的近似系数，它是函数本身的样本，叶子节点继承了该变换的近似和细节系数输出。单独的中间节点 $W_\varphi(J-1, n)$是滤波器组近似，该近似最终被滤波而变成两个叶子节点。注意，每个节点的系数都是线性展开的加权，它产生根节点 $f(n)$的一个带限"片"。因为任何这样的"片"都是一个已知尺度或小波子空间的元素(见 7.2.2 节和 7.2.3 节)，所以可以用相应的子空间来替代图 7.29(a)中的生成系数。结果是图 7.29(b)中的子空间分析树。虽然变量 W 用于表示系数和子空间，但两个量按它们的下标格式是可区分的。

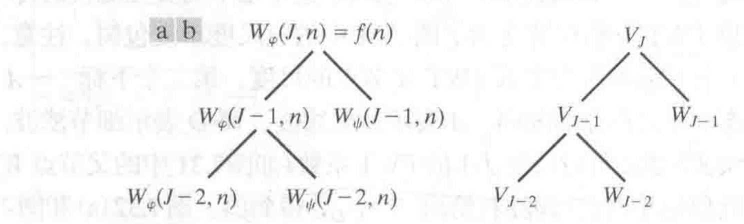


图 7.29　图 7.18 中二尺度 FWT 分析组的(a)系数树和(b)分析树

图 7.30 中进一步说明了这些概念，其中描绘了一个三尺度 FWT 分析组、分析树和相应的频谱。与图 7.18(a)不同，图 7.30(a)中的方框图标明了类似于图 7.30(b)所示的分析树和图 7.30(c)所示的频谱。这样，尽管左上滤波器和子取样器的输出是 $W_\psi(J-1, n)$，但为准确起见标为 W_{J-1}——即由 $W_\psi(J-1, n)$变换系数产生的函数子空间。这个子空间对应于相关分析树中的右上边的叶子，以及相应频谱的最右(最宽带宽)段。

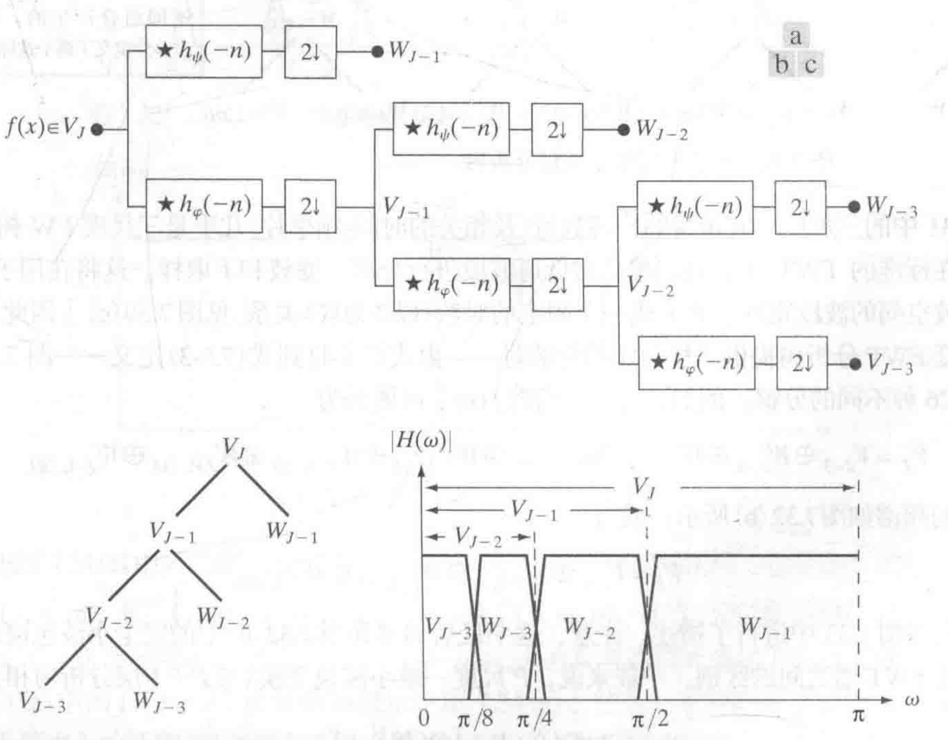


图 7.30　三尺度 FWT 滤波器组：(a)方框图；(b)分解空间树；(c)频谱分离特性

分析树提供了表示多尺度小波变换的一种紧凑且有益的方法。与基于滤波器和子取样器的相应方框图相比，分析树画起来很简单，且占用的空间较少，因此能相对容易检测有效的分解。例如，图 7.30(b)的三尺度分析树使得如下三种展开选项成为可能：

532 ~ 534

$$V_J = V_{J-1} \oplus W_{J-1} \tag{7.6-1}$$

$$V_J = V_{J-2} \oplus W_{J-2} \oplus W_{J-1} \tag{7.6-2}$$

$$V_J = V_{J-3} \oplus W_{J-3} \oplus W_{J-2} \oplus W_{J-1} \tag{7.6-3}$$

它们对应于7.4节的一尺度、二尺度和三尺度FWT分解，并可由7.2.3节的式(7.2-27)对$P=\{1,2,3\}$令$j_0=J-P$来得到。通常，P尺度FWT分析树支持唯一的P种分解。

分析树还是表示小波包的一种有效机制，小波包只不过是细节为交互滤波的传统小波变换。这样，图7.30(b)中的三尺度FWT分析树就变成了图7.31中的三尺度小波包树。注意，引入了附加的下标。双下标节点的第一个下标可视为传承FWT父节点的尺度，第二个下标——A和D的变长字符串——是从父节点到该节点的路径的编码。A表示近似滤波，而D表示细节滤波。例如，子空间$W_{J-1,DA}$是通过一个附加的细节滤波器对尺度J-1的FWT系数(如图7.31中的父节点W_{J-1})"滤波"(得到$W_{J-1,D}$)，接着用一个近似滤波器进行滤波(给出$W_{J-1,DA}$)得到的。图7.32(a)和图7.32(b)是图7.31中分析树的滤波器组和频谱分离特性。注意，图7.32(a)中的滤波器组的"自然顺序"输出，已根据图7.32(b)中的频率内容被重新排序(关于"频率排序"小波的更多内容，见习题7.25)。

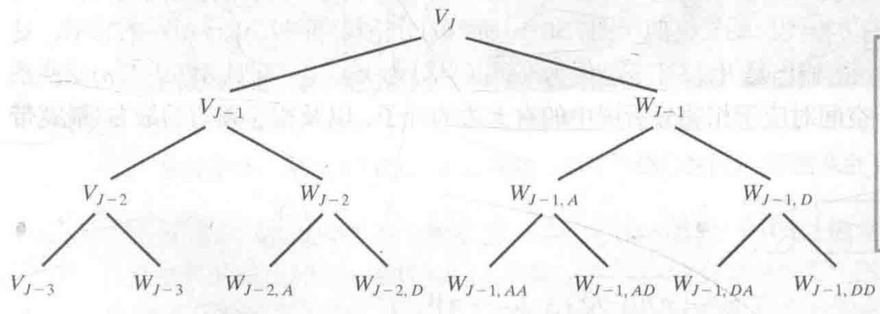


图7.31 一个三尺度小波包分析树

> 回忆可知，⊕ 表示空间的并集(就像集合的并集一样)。与图7.31相关的26种分解是由[可被组合到一起来表示树顶部的根节点(空间)的]节点(空间)的不同组合产生的。式(7.6-4)和式(7.6-5)定义了两个这样的分解。

图7.31中的三尺度小波包树的分解数量(及相关的时间-频率片)几乎是三尺度FW树的3倍。回顾可知，在标准的FWT中，我们单独对低通波段进行分离、滤波和下取样。这将在用于函数表示的尺度和小波空间的波段宽度之间形成一个固定的对数(以2为底)关系[见图7.30(c)]。因此，在图7.30(a)中的三尺度FWT分析树提供三种可能的分解时——由式(7.6-1)到式(7.6-3)定义——图7.31中的小波包树支持26种不同的分解。例如，V_J[和函数$f(n)$]可展开为

$$V_J = V_{J-3} \oplus W_{J-3} \oplus W_{J-2,A} \oplus W_{J-2,D} \oplus W_{J-1,AA} \oplus W_{J-1,AD} \oplus W_{J-1,DA} \oplus W_{J-1,DD} \tag{7.6-4}$$

它们的频谱如图7.32(b)所示；或者

$$V_J = V_{J-1} \oplus W_{J-1,A} \oplus W_{J-1,DA} \oplus W_{J-1,DD} \tag{7.6-5}$$

它们的频谱在图7.33中进行了描述。注意，这个最后的谱和图7.32(b)中的完全小波包谱或图7.30(c)中的三尺度FWT谱之间的区别。一般来说，P尺度一维小波包变换(与$P+1$层分析树相关)支持

$$D(P+1) = \left[D(P)\right]^2 + 1 \tag{7.6-6}$$

个唯一的分解，其中$D(1)=1$。随着这样的大量有效展开，基于包的变换就提供了改进被分解函数的谱的分割控制。这种控制的代价就是计算复杂性的增加[将图7.30(a)中的滤波器组与图7.32(a)进行比较]。

534 ~ 536

图 7.32　三尺度完全小波包分析树的(a)滤波器组和(b)频谱分离特性

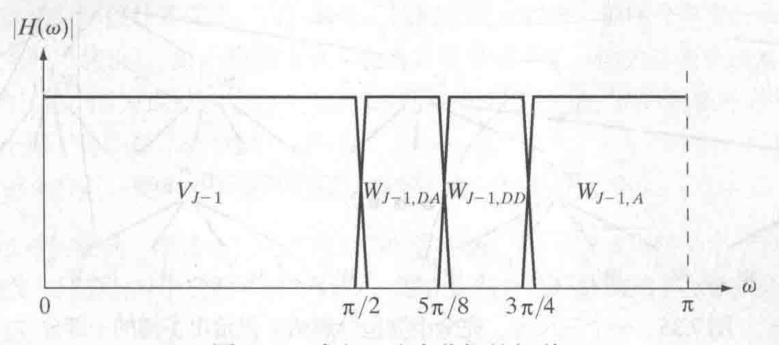


图 7.33　式(7.6-5)中分解的频谱

现在考虑图 7.24(a)中的二维、四子带滤波器组。如 7.5 节中解释的那样，它将近似 $W_\varphi(j+1, m, n)$分解成输出 $W_\varphi(j, m, n)$，$W_\psi^H(j, m, n)$，$W_\psi^V(j, m, n)$和$W_\psi^D(j, m, n)$。如在一维情况那样，在 $W_\varphi(J, m, n) = f(m, n)$的情况下，这个过程可以对尺度 $j = J-1, J-2, \cdots, J-P$ 反复迭代而生成 P 尺度变换。第一次迭代［即用图7.24(a)中的 $j+1=J$］得到的频谱示于图 7.34(a)。注意，它将频率平面分成了 4 个相等的区域。位于平面中心的低频 1/4 频带与变换系数 $W_\varphi(J-1, m, n)$及尺度空间 V_{J-1}一致(这种命名法与一维情况下一致)。然而，为适应输入的二维特性，我们现在有 3 个(而非 1 个)小波子空间。它们分别表示为 W_{J-1}^H，W_{J-1}^V 和 W_{J-1}^D，并对应于系数 $W_\psi^H(J-1,m,n)$，$W_\psi^V(J-1,m,n)$ 和 $W_\psi^D(J-1,m,n)$。图 7.34(b)显示了得到的四子带、单尺度的四分叉的 FWT 分析树。注意上标，它们将小波子空间标志及其对应的变换系数联系在一起。

535
~
536

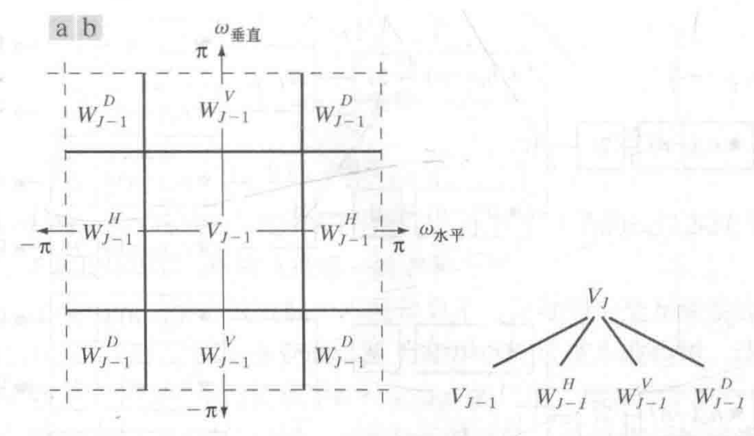


图 7.34 一个二维 FWT 的第一次分解：(a)频谱；(b)子空间分析树

图 7.35 显示了一个三尺度、二维小波包分析树的一部分。如图 7.31 中它对应的一维部分那样，传统 FWT 细节节点的每个下一代节点的第一个下标，都是其父细节节点的尺度。第二个下标(一个 A, H, V 和 D 的变长字符串)对从父节点到当前节点的路径进行编码。例如，标记为 $W_{J-1,VD}^H$ 的节点是首先用一个附加的细节/近似滤波器对尺度为 $J-1$ 的 FWT 水平细节系数(即图 7.35 中的父节点 W_{J-1}^H)进行"行/列滤波"(生成 $W_{J-1,V}^H$)，然后用一个细节/细节滤波器对它再次进行滤波（得到 $W_{J-1,VD}^H$)得到的。一个 P 尺度的二维小波包树支持唯一展开

$$D(P+1) = \left[D(P)\right]^4 + 1 \tag{7.6-7}$$

其中 $D(1)=1$。因此，图 7.35 中的三尺度树提供了 83 522 种可能的分解。在它们中间进行选择的问题是下一个例子的主题。。

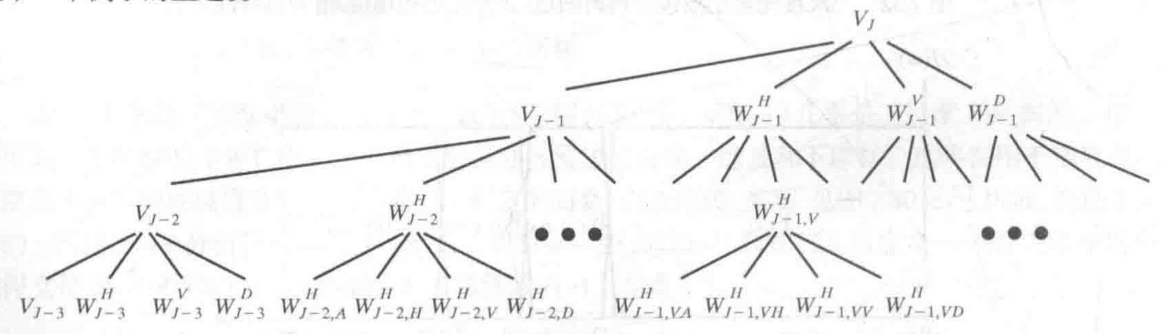


图 7.35 一个三尺度、完全小波包分解树。只给出了树的一部分

例 7.15 二维小波包分解。

如上述讨论中解释的那样，单一小波包树存在多种分解选择。实际上，可能的分解数量通常很大，以至于无法将每种分解都逐一列举出来或检验它们。因此，非常需要一种有效的算法来寻找适用于特定应用准则的最佳分解。如将要看到的那样，经典的熵和以熵为基础的代价函数在许多场合下是适用的，并且很适合应用于二叉树和四叉树搜索算法。

> 图 7.35 中的 64 个叶节点对应于图7.36(b)中 64 幅子图像的 8×8 阵列。尽管外观上相像，但它们不是方形。失真是由用于产生该结果的程序引起的(在近似子图像中尤其明显)。

考虑图7.36(a)中减少描述 400 × 400 指纹图像的数据量的问题。图像压缩将在第 8 章详细讨论。在这个例子中，我们打算选择"最好的"三尺度小波包分解作为压缩处理的起点。利用三尺度小波包树，有 83 522［见式(7.6-7)］种潜在的分解。图 7.36(b)显示了其中的一种，即一种全小波包、64 个叶子的分解，类似于图 7.35 中的分析树。注意，树的叶子对应于图 7.36(b)中被分解子图像的

536 ~ 537

8 × 8 阵列的子带。然而，以压缩为目的而使这种特定的 64 个叶子的分解达到某种程度优化的可能性相对较低。在缺乏一种合适的最佳准则时，我们既不能确认也不能否认这种分解是否最佳。

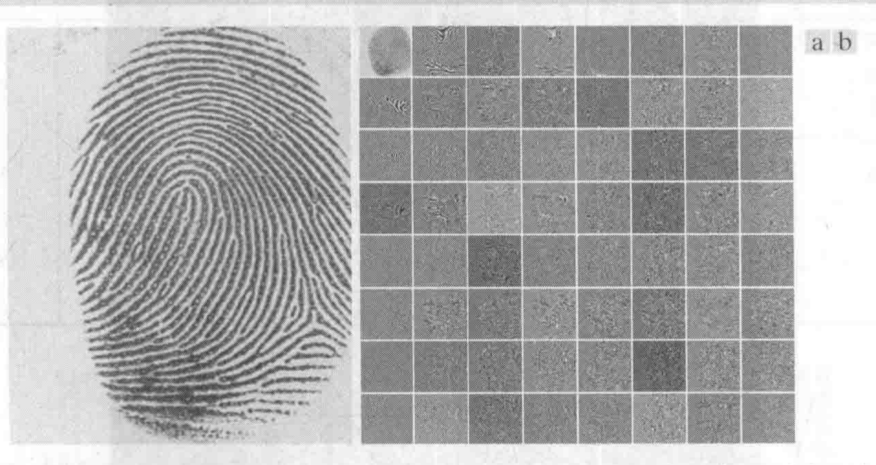


图 7.36　(a) 一幅扫描的指纹图像；(b) 该图像的三尺度、全小波包分解（原图像由美国国家标准与技术协会提供）

对图 7.36(a) 图像的压缩选择的一种合理分解准则是附加的代价函数

$$E(f)=\sum_{m,n}\left|f(m,n)\right| \tag{7.6-8}$$

该函数为二维函数 f 的能量内容提供了可能的度量。在该度量下，对于所有的 m 和 n，函数 $f(m,n)=0$ 的能量为零。另一方面，E 的高值表明函数具有很多非零值。由于大多数基于变换的压缩方案都是通过舍入或设置阈值将很小的系数变成零值来工作的，因此，从压缩的观点来看，将接近零值的数量最大化的代价函数，对于最好的分解应该是合理的准则。

> 其他可能的能量度量包括 $f(x,y)$ 的平方和、平方的对数和，等等。习题 7.27 定义了一个可能的基于熵的代价函数。

刚才提到的代价函数不仅计算简单，而且易于适应树的优化程序。最优化算法必须使用这个函数来最小化分解树中叶节点的“代价”。最小能量叶节点应该是最受欢迎的，因为这些节点具有更多接近零的数值，从而导致更大的压缩。因为式(7.6-8)的代价函数只是一个局部度量，它只用在考虑节点的有效信息上，所以很容易构造一个用于得到最小能量的有效算法。

对于分析树的每个节点，要从根节点开始逐层进行，直到到达叶子节点：

步骤 1　计算节点的能量。用 E_P（作为父节点的能量）表示，并且该节点的 4 个子节点的能量分别表示为 EA、EH、EV 和 ED。对于二维小波包分解，父节点是近似系数或细节系数的一个二维阵列；子节点分别是滤波后的近似、水平、垂直和对角线细节。

步骤 2　如果子节点的联合能量小于父节点的能量，即 $E_A+E_H+E_V+E_D<E_P$，则在分析树中包含这些子节点。如果各子节点的联合能量大于等于父节点的能量，则修剪掉这些子节点而只保留父节点。它就是优化分析树的叶子节点。

前述算法可以用于：(1) 对小波包树进行修剪，或 (2) 从零开始设计用于计算最优树的过程。在后一种情况下，不必要的兄弟节点——那些在算法的第二步中将去掉的节点后代——不用计算。图7.37 显示了优化后的分解，该分解是使用式(7.6-8)的代价函数对图 7.36(a) 中的图像应用此算法得到的。图7.38 中给出了对应的分析树。注意，图 7.36(b) 中多个原始的全小波包分解的 64 个子带（和图 7.35 中对应分析树的 64 个叶子）已被去除。另外，图 7.37 中未被分离（或进一步分解）的子图像相对比较平滑，它们由那些具有中间灰度值的像素构成。除了近似子图像，该图中的所有子图像都已被标定，所以灰度级128 指示零值系数，因此这些子图像几乎不包含能量。对这些子图像的分离应不会总体降低图像的能量。

537 ~ 539

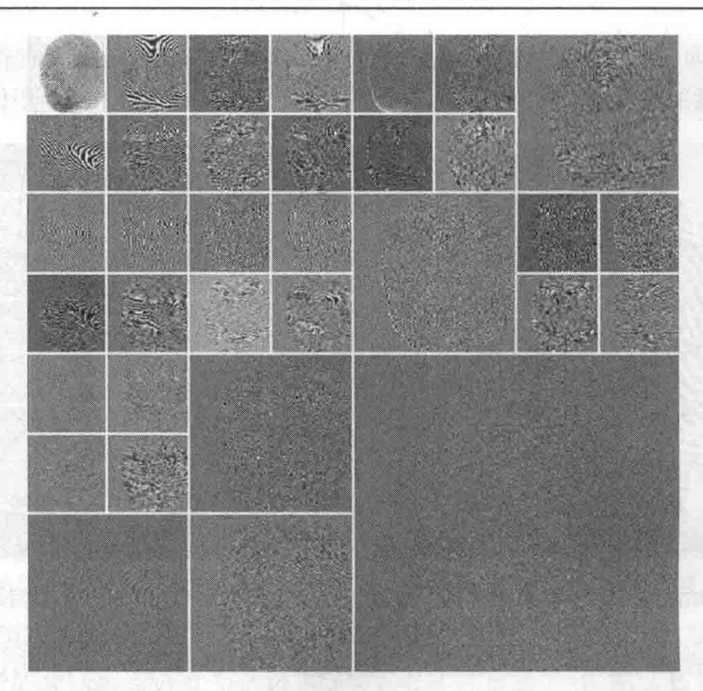


图 7.37　图 7.36(a) 中指纹的一种最佳小波包分解

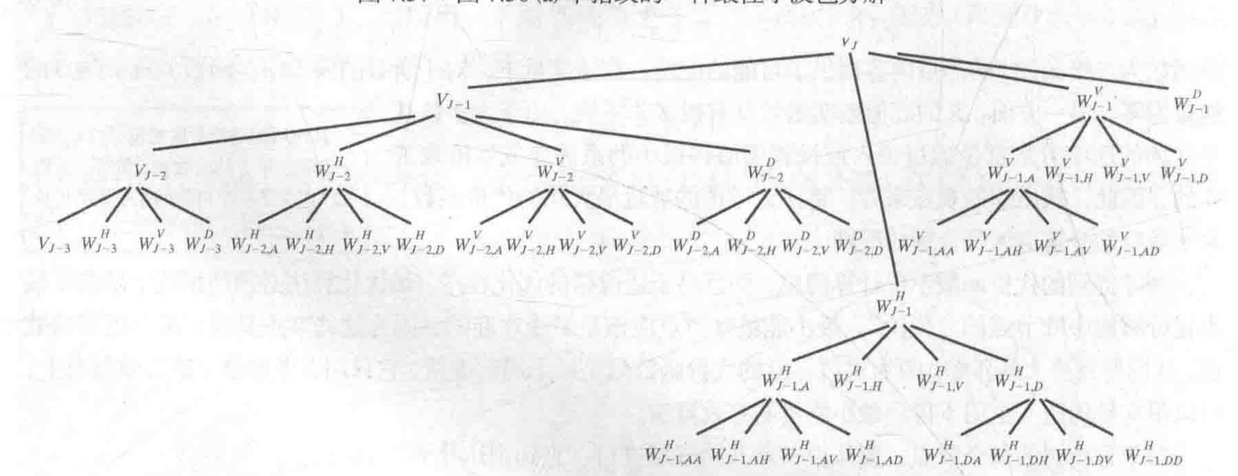


图 7.38　对图 7.37 中分解的最佳小波包分析树

前述例子是以实际问题为基础的，它可以使用小波来解决。近年来，美国联邦调查局(FBI)建立了一个巨大的指纹数据库，并为指指纹图像的数字化与压缩制定了一个基于小波的国家标准(FBI[1993])。由于使用了双正交小波，这个标准达到了典型的 15 : 1 的压缩率。小波压缩方法相对于传统 JPEG 压缩方法的优点会在下面的章节中得到检验。

例 7.15 中使用的分解滤波器和 FBI 使用的都是周知的称为 Cohen Daubechies Feauveau 双正交小波(Cohen, Daubechies and Feauveau[1992])的小波家族的组成部分。因为该小波族的尺度函数和小波函数是对称的，并且具有相近的长度，所以它们是使用最广泛的双正交小波。图 7.39(e) 到 (h) 显示了双尺度函数和小波函数。图7.39(a)到(d)是对应的分解滤波器和重建滤波器。当 $0 \leqslant n \leqslant 17$ 时，低通和高通分解滤波器 $h_0(n)$和 $h_1(n)$的系数如表 7.4 所示。相应的双正交综合滤波器的系数可使用式(7.1-11)中的 $g_0(n) = (-1)^{n+1} h_1(n)$和 $g_1(n) = (-1)^n h_0(n)$来计算。也就是说，它们是分解滤波器的交叉调制形式。注意，为使得这些滤波器的长度相同，使用了零填充，且表7.4 和图7.39 根据图7.6(a)中的子带编码和解码系统定义了这些滤波器；就 FWT 而论，有 $h_\varphi(-n) = h_0(n)$ 和 $h_\psi(-n) = h_1(n)$。

539 ~ 540

表 7.4　双正交 Cohen Daubechies Feauveau 滤波器系数（Cohen, Daubechies and Feauveau[1992]）

n	$h_0(n)$	$h_1(n)$	n	$h_0(n)$	$h_1(n)$
0	0	0	9	0.8259	0.4178
1	0.0019	0	10	0.4208	0.0404
2	−0.0019	0	11	−0.0941	−0.0787
3	−0.017	0.0144	12	−0.0773	−0.0145
4	0.0119	−0.0145	13	0.0497	0.0144
5	0.0497	−0.0787	14	0.0119	0
6	−0.0773	0.0404	15	−0.017	0
7	−0.0941	0.4178	16	−0.0019	0
8	0.4208	−0.7589	17	0.0010	0

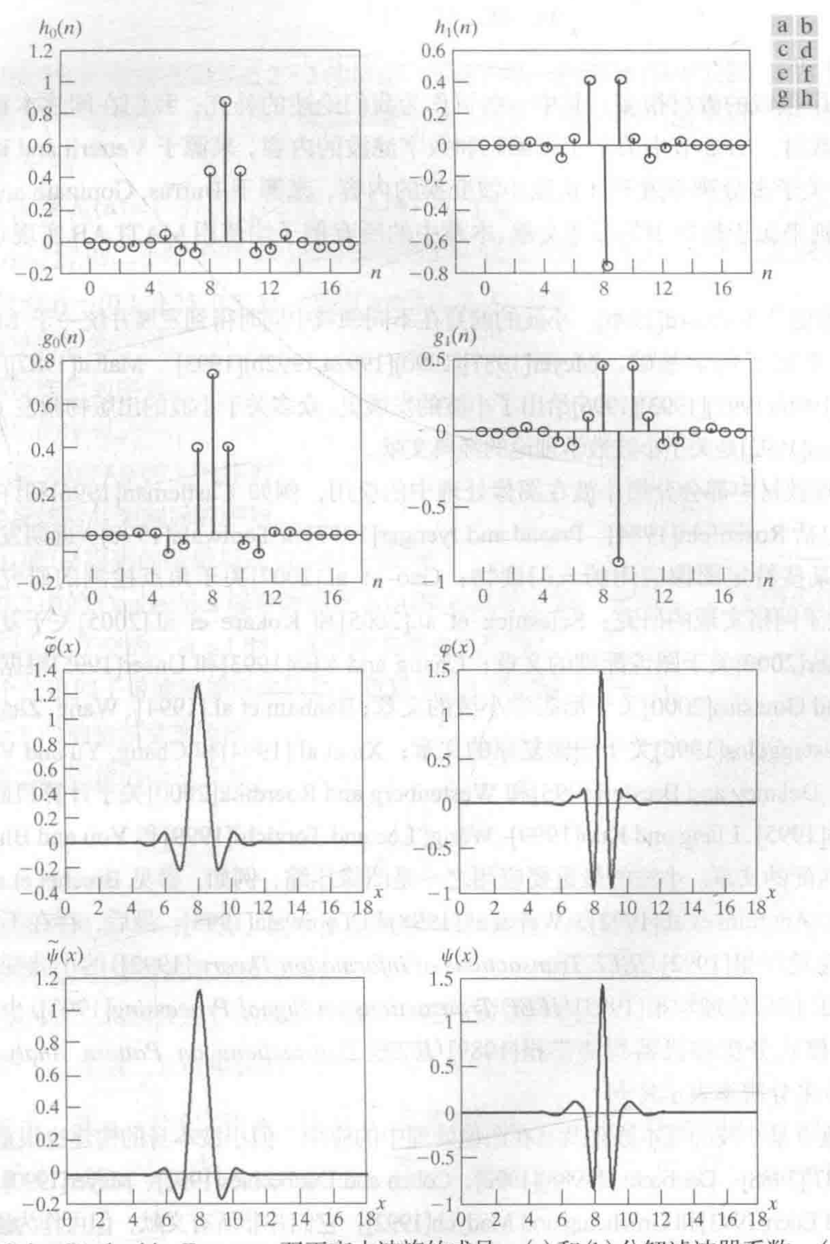


图7.39　Cohen Daubechies Feauveau 双正交小波族的成员：(a) 和 (b) 分解滤波器系数；(c) 和 (d) 重建滤波器系数；(e)~(h) 双小波和尺度函数。$0 \leqslant n \leqslant 17$ 时，$h_0(n)$ 和 $h_1(n)$ 的值见表 7.3

540 ~ 541

小结

本章为理解和深入掌握图像处理中小波与多分辨率分析的作用，奠定了牢固的数学基础。小波和小波变换是在大量图像处理课题中快速崛起的、相对较新的图像工具。由于它们与傅里叶变换具有相似性，所以第 4 章中的许多技术同样具有对应的小波域部分。成像应用中能以小波的观点来解决的问题，包括图像匹配、配准、分割、降噪、复原、增强、压缩、形态滤波和计算机断层等。由于不可能在一章中覆盖所有这些应用领域，所以根据它们的价值，只选择了部分主题进行介绍，或阐明它们的基本概念，为读者在该领域中的深入研究打下基础。第 8 章将使用小波进行图像压缩。

参考文献

小波及其应用领域的教材很多，其中一些可作为我们论述的补充；我们在阐述本章的核心内容时，依靠的也是这些教材。7.1.2 节中关于子带编码和数字滤波的内容，来源于 Vetterli and Kovacevic[1995]，7.2 节和 7.4 节中关于多分辨率展开和快速小波变换的内容，来源于 Burrus, Gopinath and Guo[1998]。本章中的其他内容则来源于教材中的参考文献。本章中的所有例子均使用 MATLAB 实现（见 Gonzalez et al. [2004]）。

小波分析的历史见 Hubbard[1998]。小波的前身在不同领域中同时得到发展并统一于 Mallat[1987]。这篇文章为小波分析奠定了数学基础。Meyer[1987][1990][1992a,1992b][1993]、Mallat[1987][1989a~c][1998]和 Daubechies[1988][1990][1992][1993][1996]给出了小波的发展史。众多关于小波的出版物激发了人们对小波的研究兴趣。Daubechies[1992]是关于小波数学理论的经典文献。

普通图像处理教材中都会介绍小波在图像处理中的应用，例如 Castleman[1996]和许多特定应用领域的书籍与学报，包括 Rosenfeld[1984]、Prasad and Iyengar[1997]和 Topiwala[1998]。近期发表的一些文章可以作为深入研究某些特定图像应用的入门读物：Gao et al.[2007]关于角点检测的研究；Olkkonen and Olkkonen[2007]关于网格实现的研究；Selesnick et al.[2005]和 Kokare et al.[2005]关于复杂小波的研究；Thévenaz and Unser[2000]关于图像配准的文章；Chang and Kuo[1993]和 Unser[1995]根据纹理进行分类的文章；Heijmans and Goutsias[2000]关于形态学小波的文章；Banham et al.[1994]、Wang, Zhang and Pan[1995]和 Banham and Kastaggelos[1996]关于图像复原的文章；Xu et al.[1994]和 Chang, Yu and Vetterli[2000]关于图像增强的文章；Delaney and Bresler[1995]和 Westenberg and Roerdink[2000]关于计算机断层摄影的文章；Lee, Sun and Chen[1995]、Liang and Kuo[1999]、Wang, Lee and Toraichi[1999]和 You and Bhattacharya [2000]关于图像描述和匹配的文章。小波的最重要应用之一是图像压缩，例如，参见 Brechet et al.[2007]、Demin Wang et al.[2006]、Antonini et al.[1992]，Wei et al.[1998]和 Topiwala[1998]。最后，存在不少关于小波的专刊，包括 IEEE 信息论学报[1992]（*IEEE Transactions of Information Theory* [1992]）的小波变换和多分辨率信号分析特刊、IEEE 信号处理学报[1993]（*IEEE Transactions on Signal Processing*[1993]）小波和信号处理特刊，以及 IEEE 模式分析和机器智能学报[1989]（*IEEE Transactions on Pattern Analysis and Machine Intelligence*[1989]）多分辨率表示特刊。

尽管本章的重点是小波的基本原理及其在图像处理中的应用，但小波本身的构建也很重要。有兴趣的读者可参阅 Battle[1987][1988]、Daubechies[1988][1992]、Cohen and Daubechies[1992]、Meyer[1990]、Mallat[1989b]、Unser, Aldroubi and Eden[1993]和 Gröchenig and Madych[1992]，它们并非所有文献，但可作为读者深入阅读的起点。读者还可参阅一些关于子带编码和滤波器组的普通文献，如 Strang and Nguyen[1996]、Vetterli and Kovacevic [1995]，以及本章中使用小波作为例子时给出的参考文献。

542 ~ 543

习题

7.1 设计一个系统，对使用图7.2(b)中的编码器生成的预测残差金字塔进行解码，并画出框图。假设编码器不会引入量化误差。

★7.2 为如下图像构建一个全填充的近似金字塔和相应的预测残差金字塔：

$$f(x,y)=\begin{bmatrix}11 & 12 & 13 & 14\\ 15 & 16 & 17 & 18\\ 19 & 20 & 21 & 22\\ 23 & 24 & 25 & 26\end{bmatrix}$$

对图 7.2(b)中的近似滤波器使用 2×2 的块进行邻域平均，并假设内插滤波器实现了像素复制。

★7.3 给定一幅 $2^J\times 2^J$ 的图像，一个 $J+1$ 级金字塔是减少还是扩展了表示该图像所需的数据量？压缩率或扩展率是多少？

7.4 包含滤波器 $h_0(n)=\{-1/\sqrt{2},1/\sqrt{2}\}$，$h_1(n)=\{-1/\sqrt{2},1/\sqrt{2}\}$，$g_0(n)=\{1/\sqrt{2},-1/\sqrt{2}\}$ 和 $g_1(n)=\{1/\sqrt{2},-1/\sqrt{2}\}$ 的二子带子带编码滤波器组是归一化正交的、双正交的？还是两者都是？

7.5 给定序列 $f(n)=\{0.1, 0.25, 0.5, 1\}$，其中 $n=0, 1, 2, 3$，计算：

(a) 符号反转后的序列。

(b) 顺序反转后的序列。

(c) 调制后的序列。

(d) 调制后并顺序反转后的序列。

(e) 顺序反转后并调制后的序列。

(f) 由(d)或(e)得到的结果与式(7.1-9)一致吗？

7.6 计算例 7.2 中的 Daubechies 综合滤波器 $g_0(n)$和 $g_1(n)$的系数。令 $m=0$，使用式(7.1-13)证明这些滤波器是归一化正交的。$m=1$ 时，这些滤波器也正交吗？

★7.7 画出一个二维四子带滤波器组解码器，重建图 7.7 中的输入 $f(m, n)$。

7.8 给出 $N=8$ 时的哈尔变换矩阵。

7.9 (a) 计算 2×2 图像的哈尔变换：

$$\boldsymbol{F}=\begin{bmatrix}4 & -2\\ 7 & 3\end{bmatrix}$$

(b) 哈尔反变换是 $\boldsymbol{F}=\boldsymbol{H}^{\mathrm{T}}\boldsymbol{T}\boldsymbol{H}$，其中 $\boldsymbol{T}$ 是 $\boldsymbol{F}$ 的哈尔变换，$\boldsymbol{H}^{\mathrm{T}}$ 是 $\boldsymbol{H}$ 的逆矩阵。证明 $\boldsymbol{H}_2^{-1}=\boldsymbol{H}_2^{\mathrm{T}}$，并用它计算(a)中结果的哈尔反变换。

7.10 对如下基计算二元组$[1,3,]^{\mathrm{T}}$的展开系数，并写出对应的展开：

★(a) 在 $\boldsymbol{R}^2$ 上，实数二元组集合的基 $\varphi_0=\left[1/\sqrt{2},1/\sqrt{2}\right]^{\mathrm{T}}$ 和 $\varphi_1=\left[1/\sqrt{2},-1/\sqrt{2}\right]^{\mathrm{T}}$。

(b) 在 $\boldsymbol{R}^2$ 上，基 $\varphi_0=[1,0]^{\mathrm{T}}$ 和 $\varphi_1=[1,1]^{\mathrm{T}}$，及其对偶 $\tilde{\varphi}_0=[1,-1]^{\mathrm{T}}$ 和 $\tilde{\varphi}_1=[0,1]^{\mathrm{T}}$。

(c) 在 $\boldsymbol{R}^2$ 上，基 $\varphi_0=[1,0]^{\mathrm{T}}$，$\varphi_1=\left[-1/2,\sqrt{3}/2\right]^{\mathrm{T}}$ 和 $\varphi_2=\left[-1/2,-\sqrt{3}/2\right]^{\mathrm{T}}$，及它们的对偶 $\tilde{\varphi}_i=2\varphi_i/3$，$i=\{0, 1, 2\}$。

(提示：须使用向量内积代替 7.2.1 节中的内积。)

543 ~ 544

7.11 证明尺度函数

$$\varphi(x)=\begin{cases}1, & 0.5\leqslant x<0.75\\ 0, & 其他\end{cases}$$

不满足多分辨率分析的第二个要求。

7.12 写一个尺度空间 V_2 的表达式作为尺度函数 $\varphi(x)$的一个函数。用式(7.2-14)中定义的哈尔尺度函数，在平移 $k=\{0,1,2,3\}$处画出哈尔 V_2 尺度函数。

★7.13 对于哈尔小波函数画出小波 $\psi_{2,2}(x)$。写出一个关于哈尔尺度函数的 $\psi_{2,2}(x)$的表达式。

7.14 假设函数 $f(x)$是哈尔尺度空间 V_2 的一个成员，即 $f(x)\in V_2$。使用式(7.2-22)将 V_2 表示为尺度空间 V_0 和任何需要的小波空间的一个函数。如果 $f(x)$在区间[0, 1)外为 0，请基于你的表达式，画出函数 $f(x)$ 的线性展开所要求的尺度函数和小波函数。

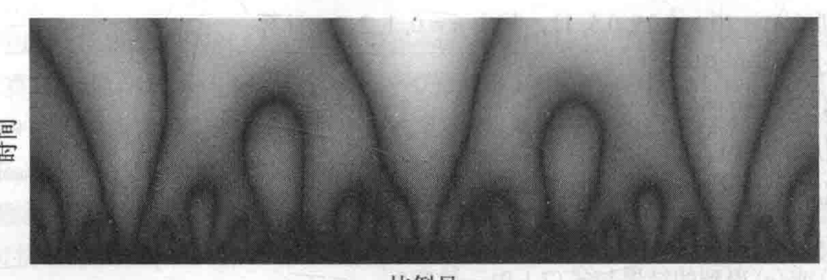
7.15 令起始尺度 $j_0=2$，计算例7.7 中所用函数的小波级数展开的前四项。将得到的展开写为所涉及尺度函数和小波函数的形式。起始尺度为 $j_0=0$ 时，得出的结果与例 7.7 的结果相比有何不同?

7.16 式(7.3-5)和式(7.3-6)中的 DWT 是起始尺度 j_0 的一个函数。

(a) 令 $j_0=1$(而不是 0)，重新计算例 7.8 中函数 $f(n)=\{1,3,0,-4\}$在区间 $0\leqslant n\leqslant 3$ 内的一维 DWT。

(b) 使用(a)的结果，根据变换值计算 $f(2)$。

★7.17 下列以一维函数为基础的连续小波变换揭示了什么?



7.18 (a) 习题7.17 的连续小波变换是由计算机生成的。以它为基础的函数首先在离散区间中取样。连续变换是什么？或者说它与函数的离散小波变换有什么区别?

★(b) 在什么条件下选择 DWT 比选择 CWT 更好？当 CWT 比 DWT 好时，能比 DWT 好几倍吗?

★7.19 画出计算习题 7.16 中的变换所需的 FWT 滤波器组。使用适当的序列标记所有的输入和输出。

7.20 M 点快速小波变换的计算复杂度为 $O(M)$，即运算次数与 M 成比例。决定比例常数的因素是什么?

7.21★ (a) 如果输入到图 7.30(a)中的三尺度 FWT 滤波器组的是哈尔尺度函数 $\varphi(n)=1, n=0,1,\cdots,7$, n 取其他值时 $\varphi(n)=0$。就哈尔变换而论，变换结果是什么?

(b) 如果输入是对应的哈尔小波函数 $\psi(n)=\{1,1,1,1,-1,-1,-1,-1\}, n=0,1,\cdots,7$，变换是什么?

(c) 什么输入序列会产生变换$\{0,0,0,B,0,0,0,0\}$，并具有非零系数 $W_\psi(1,1)=B$?

★7.22 二维快速小波变换类似于 7.2.1 节中的金字塔编码方案。它们有哪些相似性？给定图 7.10(a)中的三尺度小波变换，如何构造对应的近似金字塔？该金字塔有多少级?

7.23 计算习题7.9 中 2×2 图像的哈尔小波的二维小波变换。画出所需的滤波器组，并使用适当的阵列标记所有的输入和输出。

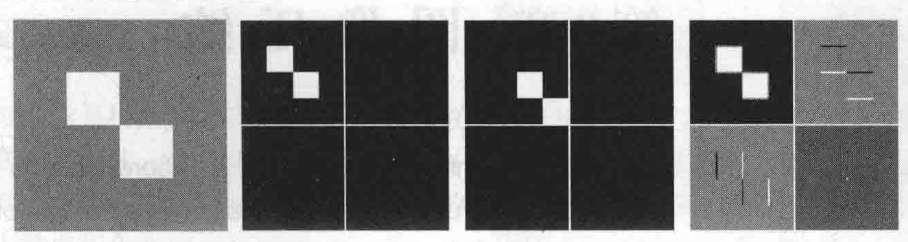
★7.24 在傅里叶域中，

$$f(x-x_0,\ y-y_0)\Leftrightarrow F(\mu,\ v)\mathrm{e}^{-2\pi(\mu x_0/M+vy_0/N)}$$

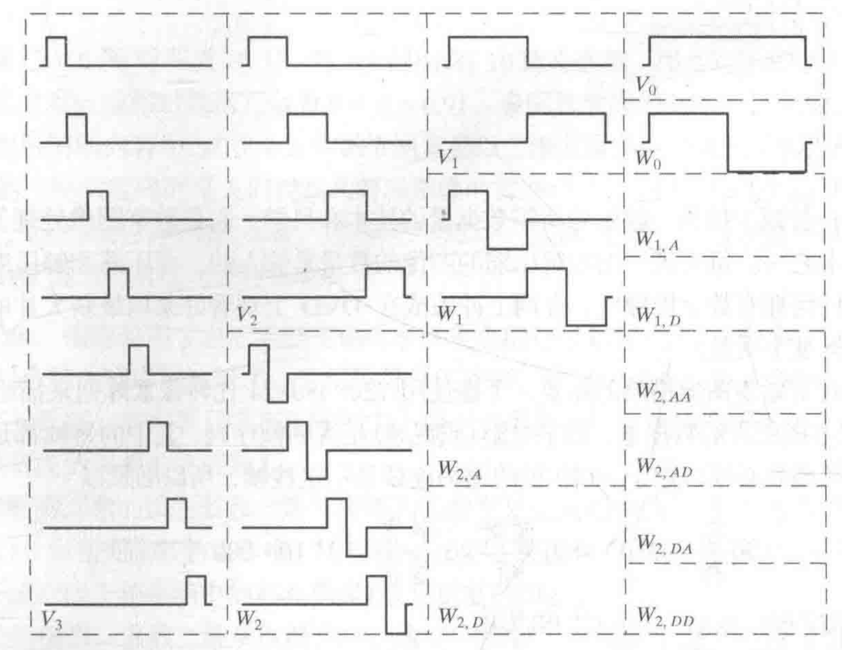
及平移并不影响$|F(\mu,v)|$的显示。使用下面的图像序列，解释小波变换的平移特性。最左边的图

544 ~ 545

像在大小为 64×64 的灰色背景的中心位置，包含两幅大小为 16×16 的白色方块。(从左数起的) 第二幅图像是第一幅图像关于哈尔小波的单尺度小波变换。第三幅图像是原图像中的方块向右下方平移 16 像素后的原始图像的小波变换结果。最后一幅图像 (最右边的图像) 是原图像向右下方移动了 1 像素后的小波变换结果。



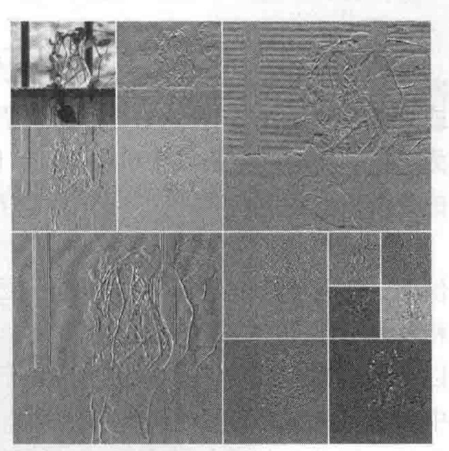
7.25 如下图表显示了一个四尺度快速小波变换的哈尔小波函数和尺度函数。大略画出完全三尺度包分解所需要的附加基函数。给出数学表达式或确定这些函数的表达式，然后根据频率内容排列基函数并解释该结果。



7.26 图 7.1 中花瓶的小波包分解显示如下：

(a) 画出对应的分解分析树，使用适当的尺度和小波空间名称标记所有节点。

(b) 画出并标记分解的频谱。



7.27 使用哈尔小波，确定函数 $f(n) = 0.5, n = 0, 1, 2, \cdots, 15$ 的最小熵包分解。使用非归一化的香农熵

$$E\left[f(n)\right] = \sum_n f^2(n)\ln\left[f^2(n)\right]$$

作为最小化判别准则。画出最佳树，使用计算的熵值标记这些节点。

第8章 图像压缩

But life is short and information endless ... Abbreviation is a necessary evil and the abbreviator's business is to make the best of a job which, Although intrinsically bad, is still better than nothing.

Aldous Huxley

引言

图像压缩是一种减少描绘一幅图像所需数据量的技术和科学，它是数字图像处理领域最有用、商业上最成功的技术之一。每天被压缩和解压缩的图像的数量是惊人的，而压缩和解压缩本身对用户而言却是无形的。任何拥有数字摄像机、在网上冲浪或在 DVD 上观看好莱坞最新大片的人，都能从本章讨论的算法和标准中受益。

为了更好地理解紧凑图像描绘的需要，考虑使用 720×480×24 比特像素阵列来描绘 2 小时的标准清晰度(SD)电视电影所需的数据量。数字电影(或视频)是视频帧序列，其中的每帧都是一幅全彩色静止图像。因为视频播放必须以近于 30 帧/秒的速率连续显示这些帧，所以必须以

$$30\frac{\text{帧}}{\text{秒}}\times(720\times480)\frac{\text{像素}}{\text{帧}}\times3\frac{\text{字节}}{\text{像素}}=31\,104\,000\ \text{字节/秒}$$

的速率访问 SD 数字视频数据。2 小时的电影由

$$31\,104\,000\frac{\text{字节}}{\text{秒}}\times(60^2)\frac{\text{秒}}{\text{小时}}\times2\text{小时}\approx2.24\times10^{11}\text{字节}$$

或 224 GB(千兆字节)的数据组成，需要 27 张 8.5 GB 的双面 DVD(假定传统的 12 cm 盘)来存储它。为了把 2 小时的电影放在单张 DVD 上，每一帧必须以平均 26.3 的系数压缩。对于高清晰度电视，它的分辨率达到 1920×1080×24 比特/图像，其压缩率应该更高。

Web 网页图像和高分辨率数字摄像机图片也要进行例行的压缩，以节省存储空间和减少传输时间。例如，住宅互联网连接发送数据的速率从传统电话线路的 56 kbps(千比特每秒)到宽带的 12 Mbps(兆比特每秒)。以该速率范围传输一幅 128×128×24 比特全彩色小图像需要的时间为 7 秒到 0.03 秒。压缩可以将传输时间减少 2~10 倍，甚至更多。以相同的方法，8 兆像素的数字摄像机在 1 GB 大小的闪存中可存储的未压缩全彩色图像的数量［约 41 幅 24 MB 图像］也以类似的数量增加。除了这些应用之外，图像压缩在许多其他领域也扮演着重要的角色，包括视频会议、遥感、文本和医学图像处理、传真传输(FAX)。应用数量的增加取决于二值图像、灰度图像和彩色图像的有效加工、存储与传输。

547~548

本章介绍数字图像压缩的理论和实践。我们将考察最常用的压缩技术，并描述使这些技术更有用的工业标准。本质上，这些内容是介绍性的，既适用于静止图像，也适用于视频应用。最后介绍数字图像水印处理，即把可见和不可见数据(如版权信息)插入到图像中的处理。

8.1 基础知识

术语数据压缩是指减少表示给定信息量所需数据量的处理。在该定义中，数据和信息是不相同的；数据是信息传递的手段。因为相同数量的信息可以用不同数量的数据表示，包含不相关或重复信息的表示称之为冗余数据。若令 b 和 b' 代表相同信息的两种表示中的比特数(或信息携带单元)，则用 b 比特表示的相对数据冗余 R 是

$$R = 1 - 1/C \tag{8.1-1}$$

式中，C 通常称为压缩率，定义为

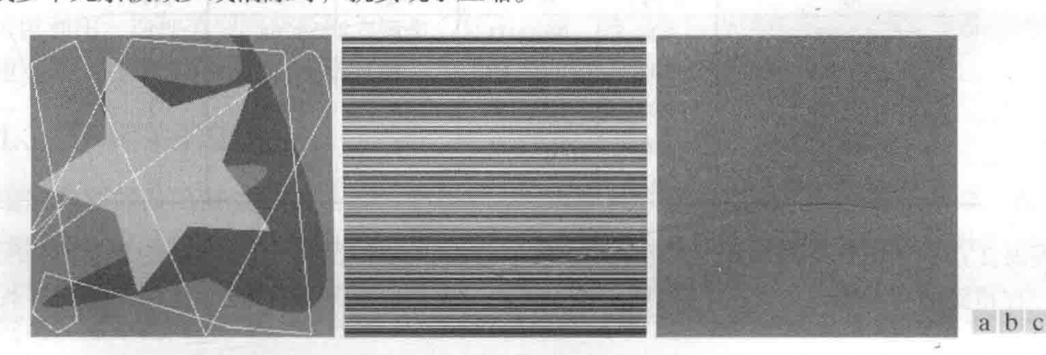
$$C = b/b' \tag{8.1-2}$$

例如，如果 $C=10$(有时写成 10:1)，较大的表示有 10 比特数据，那么对于较小表示中的每 1 比特数据，较大表示中对应的相对数据冗余为 0.9($R=0.9$)，表明其数据的 90%是冗余的。

在数字图像压缩的内容中，式(8.1-2)中的 b 通常是以二维灰度值阵列表示一幅图像所需的比特数。2.4.2 节中介绍的二维灰度阵列是人们观察和解释图像的首选格式，并以它作为判定所有其他表示的标准。然而，当它变成紧凑的图像表示时，这些格式就远不是最佳格式。二维灰度阵列受如下可被识别和利用的三种主要类型的数据冗余的影响：

1. **编码冗余**。编码是用于表示信息实体或事件集合的符号系统(字母、数字、比特和类似的符号等)。每个信息或事件被赋予一个编码符号的序列，称之为码字。每个码字中的符号数量就是该码字的长度。在多数二维灰度阵列中，用于表示灰度的 8 比特编码所包含的比特数，要比表示该灰度所需要的比特数多。
2. **空间和时间冗余**。因为多数二维灰度阵列的像素是空间相关的(即每个像素类似于或取决于相邻像素)，在相关像素的表示中，信息被没有必要地重复。在视频序列中，时间相关的像素(即类似于或取决于相邻帧中的那些像素)也是重复的信息。
3. **不相关的信息**。多数二维灰度阵列中包含有一些被人类视觉系统忽略或与用途无关的信息。从未被利用的角度看，它是冗余的。

图8.1(a)到(c)中由计算机生成的图像显示了每种基本冗余。如在接下来的三节中看到的那样，当一个或多个冗余被减少或消除时，就实现了压缩。



a b c

图 8.1　计算机产生的256 × 256 × 8 比特图像：(a)编码冗余；(b)空间冗余；(c)无关信息(设计的每幅图像都示范了一种主要的冗余，但其中也展示了其他冗余)

548～549

8.1.1 编码冗余

第 3 章在图像灰度值为随机量的假设基础上，探讨了通过直方图处理来增强图像的技术。本节将利用类似的表示方法介绍最佳信息编码。

假设我们用区间[0, L−1]内的一个离散随机变量 r_k 来表示一幅 $M\times N$ 图像的灰度，且每个 r_k 出现的概率为 $p_r(r_k)$。与 3.3 节一样，

$$p_r(r_k)=\frac{n_k}{MN},\ k=0,1,2,\cdots,L-1 \tag{8.1-3}$$

式中，L 是灰度级数，n_k 是第 k 级灰度在图像中出现的次数。如果用于表示每个 r_k 值的比特数为 $l(r_k)$，则表示每个像素所需的平均比特数为

$$L_{\text{avg}}=\sum_{k=0}^{L-1} l(r_k)p_r(r_k) \tag{8.1-4}$$

也就是说，给各个灰度级分配的码字的平均长度，可通过对用于表示每个灰度的比特数与该灰度出现的概率的乘积之和来得到。表示 $M\times N$ 图像所需的总比特数为 MNL_{avg}。如果用自然的 m 比特[①]固定长度码来表示灰度，那么式(8.1-4)的右侧将减少为 m 比特。也就是说，使用 m 来代替 $l(r_k)$时，$L_{\text{avg}}=m$。常数 m 可以提到和式之外，只剩下 $p_r(r_k)$ 在区间 $0\leqslant k\leqslant L-1$ 内的和，当然，该和为 1。

例 8.1　变长编码的简单说明。

图 8.1(a)中计算机生成的图像具有表 8.1 第二列中显示的灰度分布。如果使用 8 比特自然二进制编码(由表 8.1 中的编码 1 表示)来表示 4 种可能的灰度，则编码 1 的平均比特数 L_{avg} 为 8 比特，因为对所有 r_k 都有 $l_1(r_k)=8$ 比特。另一方面，如果使用表 8.1 中的编码 2 设计方案，则根据式(8.1-4)，编码像素的平均长度为

$$L_{\text{avg}}=0.25(2)+0.47(1)+0.25(3)+0.03(3)=1.81 \text{ 比特}$$

表示整幅图像所需要的总比特数是 $MNL_{\text{avg}}=256\times256\times1.81$ 或 118 621。由式(8.1-2)和式(8.1-1)，得到的压缩率和相应的相关冗余分别是

$$C=\frac{256\times256\times8}{118\ 621}=\frac{8}{1.81}\approx4.42 \quad 和 \quad R=1-\frac{1}{4.42}=0.774$$

这样，原始的 8 比特二维灰度阵列中，数据的 77.4%是冗余的。

表 8.1　变长编码的例子

r_k	$p_r(r_k)$	编码 1	$l_1(r_k)$	编码 2	$l_2(r_k)$
$r_{87}=87$	0.25	01010111	8	01	2
$r_{128}=128$	0.47	10000000	8	1	1
$r_{186}=186$	0.25	11000100	8	000	3
$r_{255}=255$	0.03	11111111	8	001	3
$r_k, k\neq 87,128,186,255$	0	—	8	—	0

用编码 2 达到的压缩结果来自较大可能的灰度值分配较少的比特，较小可能的灰度值分配较多的比特。在得到的变长编码中，r_{128}（图像中最大可能出现的灰度）被分配 1 比特码字 1［其长度

550 ~ 551

① 自然二进制编码是指对这种编码中将被编码的每个事件或信息(如灰度值)分配了来自一个 m 比特二进制计数序列的 2^m 种编码中的一种。

$l_2(r_{128})=1$]，r_{255}（图像中最小可能出现的灰度）被分配 3 比特码字 001 [其长度 $l_2(r_{255})=3$]。注意，可分配给图 8.1(a) 中图像的灰度的最好定长编码是自然的 2 比特计数序列{00, 01, 10, 11}，但得到的压缩率仅为 8/2 或 4 :1，比变长编码的压缩率 4.42 :1 少约 10%。

如前例所示，对事件集合(如灰度值)分配码字时，若不取全部事件概率的优势，就会出现编码冗余。用自然二进码表示一幅图像的灰度时，编码冗余几乎总是存在。其原因是大多数图像都是由规则的、在某种程度上具有可预测形态(形状)与反差的物体组成的，且这些图像被取样，所以描述的物体远大于图像元素。对大多数图像来说，自然的结果是某些灰度与其他灰度相比更可能出现(即多数图像的直方图是不均匀的)。自然二进制编码对最大和最小可能值分配相同的比特数，因此无法使式(8.1-4)最小，从而产生了编码冗余。

8.1.2　空间冗余和时间冗余

考虑图 8.1(b) 中由计算机生成的恒定灰度线的集合。在对应的二维灰度阵列中：

1. 所有 256 种灰度都是等概率的。如图 8.2 所示，图像的直方图是均匀的。
2. 因为每条线的灰度是随机选择的，在垂直方向上，每条线的像素彼此独立。
3. 因为沿每条线的像素是相同的，因此在水平方向上它们是最大相关的(完全互相依赖)。

第一个观察告诉我们，当使用传统的 8 比特灰度阵列来表示时，图8.1(b) 中的图像不能单一地使用变长编码来压缩。与图 8.1(a)（和例 8.1) 中的图像不同，这些图像的直方图是不均匀的，在这种情况下的一个定长 8 比特编码可最小化式(8.14)。观察 2 和 3 揭示了一个可被消除的有意义的空间冗余，例如，通过一个行程长度对序列来表示图 8.1(a) 中的图像，其中每个行程长度对指定一个新灰度的开始点和具有该灰度的连续像素的数量。基于行程长度的表示用(256 × 256 × 8)/[(256+256) × 8]或 128:1 的比率压缩原始的 8 比特灰度图像。原始表示的每条 256 像素的线用单一 8 比特灰度值和行程表示中的长度 256 来替代。

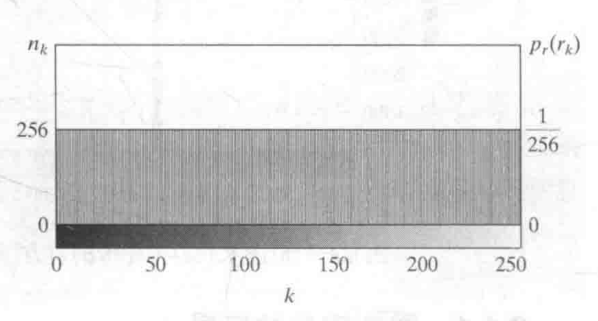


图 8.2　图 8.1(b) 中图像的灰度直方图

在多数图像中，像素是空间(在 x 和 y 方向)和时间相关的(当该图像是视频序列的一部分时)。因为多数像素灰度可根据相邻像素灰度进行合理的预测，所以单个像素携带的信息较少。在这种意义上，一个像素可由其相邻像素推断出来，因此其视觉贡献的大多数就是冗余的。为减少空间与时间相关像素涉及的冗余，二维灰度阵列必须变换为更有效但通常不可见的表示。例如，行程长度或相邻像素之间的差可供利用。这种类型的变换称为映射。如果原始二维灰度阵列的像素可以根据变换后的数据集合无误地重建，则称这个映射是可逆映射，否则称这个映射是不可逆映射。

8.1.3　不相关的信息

压缩数据集的最简方法之一是，从集合中消除多余的数据。在数字图像压缩方面，被人类视觉系统忽略的信息或与图像预期的应用无关的信息显然都是要删除的对象。这样，图 8.1(c) 中由计算机生成的图像，就可单独由其平均灰度(单一的 8 比特值)来表示，因为它是以单一灰度场的形式出现的。原始的 256×256×8 比特灰度阵列被减少到单个字节；得到的压缩率为 256×256×8/8 或 65 536 : 1。当然，原始 256×256×8 比特图像必须被重建以供观察或对它进行分析，但重建图像的质量会有较少的或不被察觉的下降。

551 ~ 553

图8.3(a)显示了图8.1(c)中的图像的直方图。注意，实际存在一些灰度值(灰度125到131)，人的视觉系统平均了这些灰度，并仅感知平均值，而忽略了这种情况下灰度的较小变化。图 8.3(b)是图 8.1(c)中图像经直方图均衡处理后的结果，它使得灰度变化可见，并显示了两个先前不可检测的恒定灰度区域，一个是垂直的，另一个是水平的。如果图 8.1(c)中的图像用其平均灰度值单独表示，这个“不可见的”结构(即两个恒定灰度区域)及其周围的随机灰度变化(真实信息)将会丢失。这些信息是否保留由应用决定。如果该信息很重要，如它可能应用于医学(如数字X射线档案)中，就不应该遗漏；否则，这种信息就是冗余的，为确保压缩性能可将它排除在外。

我们注意到，这里考察的冗余度与8.1.1节和8.1.2节中讨论的冗余度的根本不同，给出了本节的结论。消除冗余是可能的，因为这种信息本身对于正常的视觉处理和/或期望的图像用途并不是本质的。由于去除这种信息会导致定量信息的损失，因此这种信息的去除通常称为量化。这一术语与该词的标准用法一致，它通常意味着将较宽范围的输入值映射为有限数量的输出值(见2.4节)。因为信息损失了，所以量化是一种不可逆的操作。

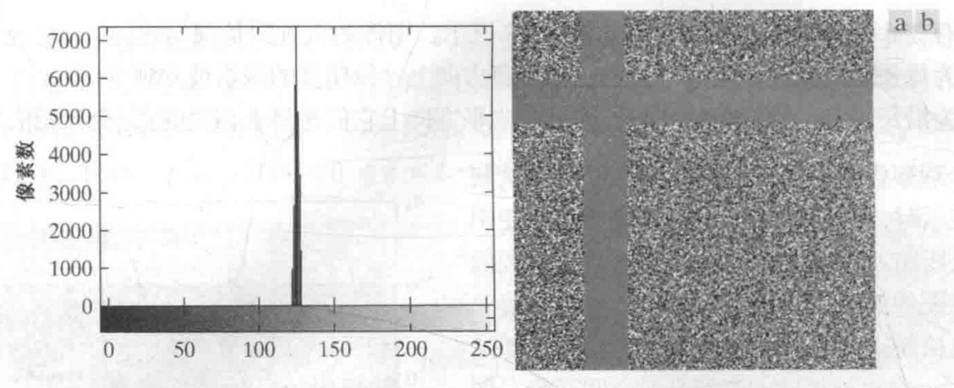


图 8.3 (a)图 8.1(c)中图像的直方图；(b)该图像经直方图均衡处理后的形式

8.1.4 图像信息的度量

前几节介绍了减少用于表示一幅图像的数据量的几种方法。自然地，出现的问题是：表示一幅图像中的信息实际上需要多少比特？也就是说，存在不丢失信息的充分描述一幅图像的最小数据量吗？信息论提供了回答这个问题和相关问题的数学框架。它的基本前提是，信息的产生可用一个概率过程建模，该过程可以用一种与直觉一致的方式加以度量。根据这一推测，我们可以说概率为$P(E)$的随机事件E包含

> 关于信息论和概率论的简要回顾，请参阅本书的网站。

$$I(E) = \log\frac{1}{P(E)} = -\log P(E) \tag{8.1-5}$$

单位的信息。如果$P(E)=1$(即事件总会发生)，则$I(E)=0$，并认为它没有信息。因为相关的该事件没有不确定性，所以在事件发生的通信过程中不会传递任何信息［如果$P(E)=1$，该事件总会发生］。

式(8.1-5)中对数的底决定了度量信息所用的单位。如果使用以m为底的对数，则这种度量称为m元单位。如果将底选择为2，则信息的单位是比特。注意，如果$P(E)=1/2$，那么$I(E)=-\log_2 1/2$或1比特。也就是说，当两个出现概率相等的事件之一发生时，传达的信息量是1比特。一个简单的例子就是投掷一枚硬币及其传递的结果。

553 ~ 554

从一个可能事件的离散集合$\{a_1, a_2, \cdots, a_J\}$，给定一个统计独立随机事件的信源，与该集合相联系的概率为$\{P(a_1), P(a_2), \cdots, P(a_J)\}$，则每个信源输出的平均信息称为该信源的熵，即

$$H = -\sum_{j=1}^{J} P(a_j)\log P(a_j) \tag{8.1-6}$$

式中，a_j 称为信源符号。因为它们是统计独立的，所以信源本身称为零记忆信源。

把一幅图像考虑为一个虚构零记忆“灰度信源”的输出时，我们可通过观察图像的直方图来估计该信源的符号概率。这时，灰度信源的熵变为

$$\tilde{H} = -\sum_{k=0}^{L-1} p_r(r_k)\log_2 p_r(r_k) \tag{8.1-7}$$

> 式(8.1-6)适用于有着 J 个信源符号的零记忆信源；式(8.1-7)使用一幅图像中的 L–1 个灰度值的概率估计。

式中，变量 L, r_k 和 $p_r(r_k)$ 如 8.1.1 节和 3.3 节中的定义。因为使用以 2 为底的对数，式(8.1-7)是以比特(bit)度量的虚构灰度信源的每个灰度输出的平均信息。它不可能对虚构信源的灰度值以比 $\tilde{H}$ 比特/像素少的码(取样图像)来编码。

例 8.2　图像的熵的估计。

图 8.1(a)所示图像的熵，可通过把表 8.1 中的灰度概率代入式(8.1-7)中来估计：

$$\begin{aligned}\tilde{H} &= -\left[0.25\log_2 0.25 + 0.47\log_2 0.47 + 0.25\log_2 0.25 + 0.03\log_2 0.03\right]\\ &\approx -\left[0.25(-2) + 0.47(-1.09) + 0.25(-2) + 0.03(-5.06)\right]\\ &\approx 1.6614 \text{ 比特/像素}\end{aligned}$$

采用类似的方法，图 8.1(b)和(c)中图像的熵可以分别表示为 8 比特/像素和 1.566 比特/像素。注意，图 8.1(a)中的图像似乎具有更多的视觉信息，但计算出的熵却几乎是最低的——1.66 比特/像素。图8.1(b)中图像的熵几乎是图8.1(a)中图像的熵的 5 倍，但似乎有相同(或更少)的视觉信息；图8.1(c)中似乎有很少的信息或没有信息，却具有与图 8.1(a)中图像几乎相同的熵。因而明显的结论是，一幅图像中熵的数量和信息与直觉相差甚远。

香农第一定理

回顾可知，例 8.1 中的变长编码仅用 1.81 比特/像素就可表示图 8.1(a)中图像的灰度。虽然它比例 8.2 中的估计熵 1.6614 比特/像素要高，但香农第一定理［也称为*无噪声编码定理*(Shannon [1948])］告诉我们，图 8.1(a)中的图像可用不多于 1.6614 比特/像素的熵来表示。为了用一般的方法来证明它，香农使用单个码字(而不是每个信源符号一个码字) 考察了 n 个连续信源符号表示的组合，并证明了

$$\lim_{n\to\infty}\left[\frac{L_{\text{avg},\,n}}{n}\right] = H \tag{8.1-8}$$

式中，$L_{\text{avg},n}$ 是表示所有 n 个符号组所需编码符号的平均数。在证明中，他将一个零记忆信源的第 n 个扩展定义为一个假设的信源，该假设信源使用原始信源符号产生 n 个符号块；并将式(8.1-4)应用到表示 n 个符号块[①]的码字来计算 $L_{\text{avg},n}$。式(8.1-8)告诉我们，$L_{\text{avg},n}/n$ 可用单一符号源的无限扩展编码任意接近 H。也就是说，用每信源符号 H 信息单位的平均来表示零记忆信源的输出是可能的。

① 第 n 个扩展的输出是一个 n 元符号组，这些符号来自底层的单一符号信源。它被视为一个块随机变量，在该变量中，每个 n 元组的概率是各个符号的概率的积。因而，第 n 个扩展的熵，是单一符号信源的熵的 n 倍。

554~555

如果现在我们回到图像是产生其灰度信源的一个“取样”这一概念，那么n个信源符号的一个块对应于n个邻近像素组。为了对n个像素块构建一个变长编码，必须计算块的相关的频数。但是，具有256个灰度值的假设的灰度信源的第n个扩展有256^n个可能的n像素块。甚至在$n=2$的简单情况下，也需要产生65 536个元素直方图和多达65 536个变长码字。对于$n=3$，需要多达16 777 216个码字。因此，即使对于较小的n值，在实践中，计算的复杂性也限制了扩展编码方法的使用价值。

最后，我们注意到，在对统计独立的像素直接编码时，尽管式(8.1-7)提供了可以达到的压缩的下界，但当一幅图像的像素相关时，它就会失败。相关像素块可用比公式预测的每像素平均比特数低的码率编码。与其用不相关描述符的信源扩展(类似灰度行程)，不如选择无扩展的普通方法。这是在8.1.2节中用于压缩图8.1(b)的方法。当信息源的输出依赖于前面的有限数量的输出时，那么该信源就称为马尔可夫信源或有限记忆信源。

8.1.5 保真度准则

在8.1.3节中，我们注意到去除“与视觉不相关”的信息会导致选择真实的或一定数量的图像信息。因为信息的丢失，因此需要一种量化这种丢失的本质的方法。两类准则可用于这样的评估：(1)客观保真度准则；(2)主观保真度准则。

当信息损失可以表示为压缩处理的输入和输出的数学函数时，则称其是以客观保真度准则为基础的。一个例子是两幅图像间的均方根(rms)误差。令$f(x,y)$是输入图像，并令$\hat{f}(x,y)$是$f(x,y)$的近似，它来自对输入先压缩后解压缩的结果。对x和y的所有值，$f(x,y)$和$\hat{f}(x,y)$之间的误差$e(x,y)$为

$$e(x,y)=\hat{f}(x,y)-f(x,y) \tag{8.1-9}$$

因此，两幅图像间的总误差为

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}(x,y)-f(x,y)\right]$$

其中，图像的大小为$M\times N$。而$f(x,y)$和$\hat{f}(x,y)$之间的均方根误差e_{rms}是在$M\times N$阵列上平均误差的平方的平方根，或写为

$$e_{\text{rms}}=\left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}(x,y)-f(x,y)\right]^2\right]^{1/2} \tag{8.1-10}$$

如果［简单地重排式(8.1-9)中的各项］认为$\hat{f}(x,y)$是原始图像$f(x,y)$和一个误差或“噪声”信号$e(x,y)$的和，则用SNR_{ms}表示的输出图像的均方信噪比可像在5.8节中那样定义为

$$\text{SNR}_{\text{ms}}=\frac{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\hat{f}(x,y)^2}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left[\hat{f}(x,y)-f(x,y)\right]^2} \tag{8.1-11}$$

表示为SNR_{ms}的信噪比的均方根值，可通过求式(8.1-11)的平方根得到。

尽管客观保真度准则提供了评估信息损失的一种简便方法，但解压缩后的图像最终还是由人来观察的。因此，使用人的主观评估来衡量图像的质量通常更为适当。主观评估是通过向观察者显示解压缩的图像，并将他们的评估结果进行平均得到的。评估可使用一个绝对等级尺度或借助于$f(x,y)$和$\hat{f}(x,y)$的并排比较来获得。表8.2显示了一种可能的绝对等级尺度。并排比较可通过使用像

556 ~ 557

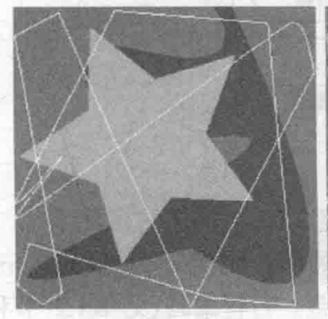
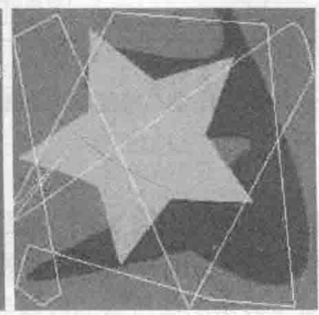
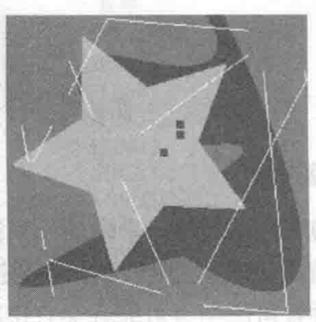
{−3, −2, −1, 0, 1, 2, 3}这样的一个等级分别表示主观评估｛非常恶劣、恶劣、稍坏、普通、稍好、更好、非常好｝。不管使用何种形式，这些评估均基于主观保真度准则。

例 8.3 图像质量比较。

图 8.4 显示了图 8.1(a)中图像的三种不同的近似。将图 8.1(a)中的图像作为 $f(x,y)$，并将图 8.4(a)中的图像作为 $\hat{f}(x,y)$，用式(8.1-10)计算的均方根误差分别是 5.17, 15.67 和 14.17 个灰度级。根据均方根误差（客观保真度准则），图 8.4 中的三幅图像按质量递减的顺序排列为{(a), (c), (b)}。

表 8.2 电视配置研究组织的等级尺度(Frendendall and Behrend)

值	等 级	描 述
1	优秀	一幅极高质量的图像，和您希望的一样好
2	好	一幅高质量的图像，提供赏心悦目的观看效果。干扰不讨厌
3	可通过的	一幅质量可接受的图像。干扰不讨厌
4	边缘的	一幅质量较差的图像，您希望能改进它，干扰有些讨厌
5	差	一幅质量非常差的图像，但还可以观看，明显存在讨厌的干扰
6	不能用	一幅坏到不能观看的图像

a b c

图 8.4 图 8.1(a)中图像的三个近似

图8.4(a)和(b)是已被压缩并随后被重建的典型图像。两幅图像都保留了原图像的基本信息，如目标的空间位置和灰度特点。它们的均方根误差粗略地对应于感觉的质量。实际上与原图像一样好的图 8.4(a)具有最低的均方根误差，而图8.4(b)具有更多的误差，但在物体间的边界处存在显而易见的退化。这恰好如预期的那样。

图8.4(c)是人工生成的图像，它给出了客观保真度准则的限制。注意，在右上象限中，图像失去了几条重要线条的大部分(即可视信息)，并出现了几个暗色的小方块(即人为缺陷)。图像的视觉内容被误解了，且的确不像图(b)中的图像那样准确，但它有着更小的均方根误差——14.17 灰度级对 15.67 灰度级。使用表 8.2 对这三幅图像进行主观评估，可能会产生对图(a)为“优秀”的等级，对图(b)为“可通过的”或“边缘的”的等级，对图(c)是“差”或“不能用”的等级。另一方面，用均方根误差来度量时，(c)却排在(b)的前面。

8.1.6 图像压缩模型

如图 8.5 所示，图像压缩系统是由两个不同的功能部分组成的：一个编码器和一个解码器。编码器执行压缩操作，解码器执行解压缩操作。两种操作可用软件执行，如在 Web 浏览器和许多商业图像编辑程序中那样，或者使用硬件和固件相结合的形式执行，如商业 DVD 播放器。codec 是一个具有编码和解码能力的装置或程序，它有编码和解码的能力。

557 ~ 558

图像 $f(x,\cdots)$ 被输入到编码器中，这个编码器创建该输入的压缩表示。把这一表示存储起来以备后续应用，或为传输而存储，以便远程应用。当压缩后的表示送入其互补的解码器中时，就会产生重建的输出图像 $\hat{f}(x,\cdots)$。在静止图像应用中，编码的输入和解码器的输出分别是 $f(x,y)$ 和 $\hat{f}(x,y)$；在视频应用中，它们分别是 $f(x, y, t)$ 和 $\hat{f}(x,y,t)$，其中离散参数 t 规定为时间。通常，$\hat{f}(x,\cdots)$ 可能是也可能不是 $f(x,y)$ 的精确副本。若是 $f(x,y)$ 的精确副本，则压缩系统被称为无误差的、无损的或信息保持的压缩系统；若不是 $f(x,y)$ 的精确副本，则重建的输出图像就会失真，且压缩系统称为有损压缩系统。

这里，$f(x,\cdots)$ 用于表示 $f(x,y)$ 和 $f(x,y,t)$。

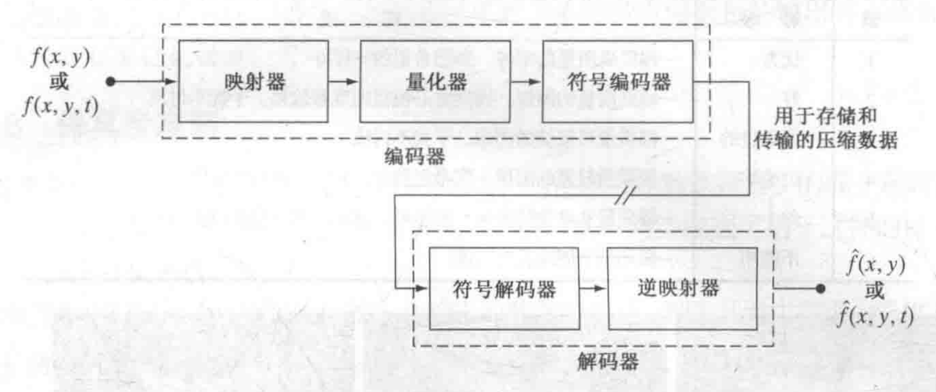


图 8.5 一个通用图像压缩系统的功能方框图

编码或压缩过程

图 8.5 中的编码器通过一系列的三个独立操作，去除 8.1.1 节至 8.1.3 节描述的冗余形式。在编码处理的第一个阶段，映射器把 $f(x,\cdots)$ 变换为降低空间和时间冗余的形式(通常不可见)。这一操作通常是可逆的，并且可能会也可能不会直接减少表示图像所需的数据量。行程编码(见 8.1.2 节和 8.2.5 节)就是一个映射的例子，该映射通常在编码处理的第一步中就会得到压缩。把一幅图像映射为一组不相关的变换系数(见 8.2.8 节)是相关情况的一个例子(为实现压缩，必须对系数进行进一步处理)。在视频应用中，映射器使用前面的(在某些将来的情况下)视频帧来帮助去除时间冗余。

图8.5中的量化器根据预设的保真度准则降低映射器输出的精度，目的是排除压缩表示的无关信息。如 8.1.3 节中解释的那样，这一操作不可逆。当我们希望进行无误差压缩时，这一步必须略去。在视频应用中，通常需要度量编码输出的比特率(比特/秒)，并调整量化器的操作，以保持预设的平均输出比特率。这样，输出的视觉质量就可根据图像内容逐帧变化。

在第三阶段，即信源编码处理的最后阶段，图 8.5 中的符号编码器生成一个定长编码或变长编码来表示量化器的输出，并根据该编码来变换输出。大多数情况下会使用变长编码。最短的码字赋予出现频率最高的量化器输出值，以最小化编码冗余。这种操作是可逆的。这一操作完成后，输入图像就完成了 8.1.1 节到 8.1.3 节中所描述的三种冗余去除。

解码或解压缩过程

图 8.5 中的解码器仅包含两个部分：一个符号解码器和一个反映射器。它们以相反的顺序执行编码器的符号编码器和映射器的反操作。因为量化导致了不可逆的信息损失，所以反量化器模块未包含在通常的解码器模型中。在视频应用中，解码后的输出帧保留在内部帧存储器中(未显示)，以重新插入在编码器中去除的时间冗余。

558~560

8.1.7 图像格式、容器和压缩标准

在数字图像处理领域，图像文件格式是组织和存储图像数据的标准方法。它定义了所用的数据排列方式和压缩类型——如果有的话。图像容器类似于文件格式，但处理多种类型的图像数据。另一方面，图像压缩标准对压缩和解压缩图像定义了过程，也就是定义减少表示一幅图像所需数据量的过程。这些标准是图像压缩技术被普遍接受的基础。

图 8.6 列出了当前使用的大多数重要的图像压缩标准、文件格式和容器，并按被处理图像的类型进行了归类。加粗的条目是由国际标准化组织(ISO)、国际电工委员会(IEC)和国际电信联盟［ITU-T，一个联合国组织，从前称为国际电报电话咨询委员会(CCITT)］批准的国际标准。还包含两个视频压缩标准，即由电影和电视工程师协会(SMPTE)批准的 VC-1 和由中国原信息产业部(MII)批准的 AVS。注意，图 8.6 中的灰色条目不是国际标准化组织支持的标准。

表 8.3 和表 8.4 总结了图 8.6 中列出的标准、格式和容器。也指出了负责组织、应用目标和关键的压缩方法。压缩方法本身是下一节的主题。两个表中以方括号的形式提前引用了 8.2 节中的相关小节。

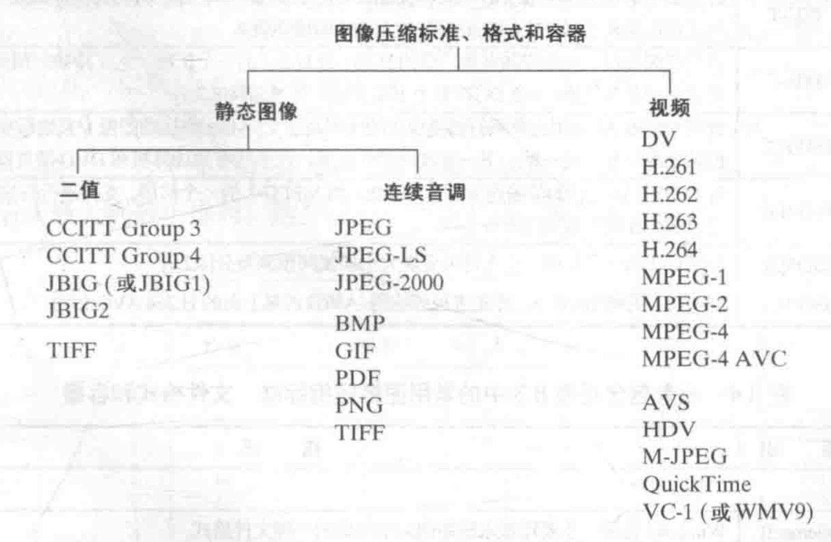


图 8.6 一些常用的图像压缩标准、文件格式和容器。国际支持的条目显示为黑色，其他条目显示为灰色

表 8.3 国际支持的压缩标准。方括号中的数字指的是本章中的小节号

名 称	组 织	描 述
二值静止图像		
CCITT 3 组	ITU-T	为在电话线上传输二值文件的传真(FAX)设计的方法。支持一维和二维行程编码[8.2.5]及霍夫曼编码[8.2.1]
CCITT 4 组	ITU-T	CCITT 3 组标准的一个精简和革新的版本，仅支持二维行程编码
JBIG 或 JBIG1	ISO/IEC/ITU-T	用于渐进、无损压缩二值图像的一个联合二值图像专家组标准。高达 6 比特/像素的连续色调图像可在一个位面的基础上编码[8.2.7]。使用了上下文敏感的算术编码[8.2.3]，且初始低分辨率的图像版本可使用附加的压缩数据逐步增强
JBIG2	ISO/IEC/ITU-T	遵循 JBIG1 的，针对桌面、互联网和 FAX 应用的二值图像的标准。使用的压缩方法是基于内容的：对文本和半色调区域使用基于字典的方法[8.2.6]，而对其他图像内容使用霍夫曼编码[8.2.1]或算术编码[8.2.3]。它可以是有损的或无损的
连续色调静止图像		
JPEG	ISO/IEC/ITU-T	针对图片质量的图像的联合图片专家组标准。其有损基准编码系统(最通用的实现)在8 × 8图像块上[8.2.8]用量化的离散余弦变换(DCT)、霍夫曼编码[8.2.1]和行程编码[8.2.5]。它是在互联网上压缩图像的最普遍的方法之一

560
~
561

(续表)

名 称	组 织	描 述
连续色调静止图像		
JPEG-LS	ISO/IEC/ITU-T	针对连续色调图像的基于自适应预测[8.2.9]、上下文模型[8.2.3]和 Golomb 编码[8.2.2]的无损到接近无损的标准
JPEG-2000	ISO/IEC/ITU-T	遵循 JPEG 的用于提高图片质量图像的压缩率的标准。它使用算术编码[8.2.3]和量化离散小波变换(DWT)[8.2.10]。压缩可以是有损的或无损的
视频		
DV	IEC	数字视频。适用于家庭和半专业视频制作应用，以及类似电子新闻采集和便携式摄像机等设备的视频标准。它使用一种类似于 JPEG 的基于 DCT 的方法[8.2.8]对简单编辑目的单独地压缩图像帧
H.261	ITU-T	针对 ISDN 线路(综合业务数字网)的一种双向视频会议标准。它分别支持非隔行扫描的 352×288 和 176×144 分辨率图像［称为 CIF(通用中间格式)和 QCIF(四分之一 CIF)］。采用类似于 JPEG 的基于 DCT 的压缩方法[8.2.8]，使用帧间预测差分[8.2.9]来减少时间冗余。基于块的技术用于帧间的运动补偿
H.262	ITU-T	见下面的 MPEG-2
H.263	ITU-T	针对具有附加分辨率 SQCIF (Sub-Quarter CIF 128×96)、4CIF(704×576)和 16CIF(1408×512)的普通电话调制解调器(如 28.8 Kb/s)设计的 H.261 的增强版本
H.264	ITU-T	针对视频会议、网络流和电视广播的 H.261 至 H.263 的一个扩展。它支持帧间预测差分[8.2.9]、可变块大小的整数变换(不是 DCT)和上下文自适应算术编码[8.2.3]
MPEG-1	ISO/IEC	针对 CD-ROM 应用的非隔行扫描视频的速率高达 1.5 Mb/s 的运动图像专家组标准。它类似于 H.261，但帧预测可基于前一帧、下一帧或两者的内插。几乎所有的计算机和 DVD 播放器都支持该标准
MPEG-2	ISO/IEC	针对 DVD 所设计的传输速率高达 15 Mb/s 的 MPEG-1 的一个扩展。支持隔行扫描视频和 HDTV。它是目前为止最为成功的视频标准
MPEG-4	ISO/IET	MPEG-2 的一个扩展，它支持可变块大小和帧间预测差分[8.2.9]
MPEG-4AVC	ISO/IET	MPEG-4 的第 10 部分，即先进视频编码(AVC)。与上面的 H.264 AVC 相同

表 8.4 没有包含在表 8.3 中的常用图像压缩标准、文件格式和容器

名 称	组 织	描 述
连续色调静止图像		
BMP	Microsoft	Windows 位图。主要用做未压缩的简单图像的一种文件格式
GIF	CompuServe	图形交换格式。对 1 到 8 比特图像使用无损 LZW 编码[8.2.4]的文件格式。它常用于为万维网制作小动画和低分辨率影片
PDF	Adobe Systems	便携文件格式。针对与设备和分辨率无关的方法表示二维文档的一种格式。它可以像 JPEG、JPEG-2000、CCITT 和其他压缩图像的容器来运行。某些 PDF 版本已成为 ISO 标准
PNG	万维网联盟(W3C)	便携网络图形。通过对每个像素值与基于过去像素的预测值间的差进行编码，是全透明(高达 48 比特/像素)无损地压缩全彩色图像[8.2.9]的文件标准。
TIFF	Aldus	标记图像格式。一种灵活的文件格式，支持多种图像压缩标准，包括 JPEG、JPEG-LS、JPEG-2000、JBIG2 和其他标准
视频		
AVS	MII	音视频标准。类似于 H.264，但使用指数 Golomb 编码[8.2.2]。中国开发的标准
HDV	公司联盟	高清晰度视频。针对高清电视的 DV 的一个扩展，它使用类似于 MPEG-2 的压缩，包括使用预测差分方法[8.2.9]去除时间冗余
M-JPEG	多家公司	活动 JPEG。一种压缩格式，在这种格式中，每一帧都独立地用 JPEG 来压缩
Quick-Time	Apple Computer	支持 DV、H.261、H.262、H.264、MPEG-1、MPEG-2 和 MPEG-4 他视频压缩格式的一种媒体容器
VC-1	SMPTE	互联网上最为通用的视频格式
WMV9	Microsoft	适用于高清和蓝光高分辨率 DVD 的一种格式。它类似于 H.264/AVS，使用一个可变块大小的整数 DCT[8.2.8, 8.2.9]和依赖于上下文的变长码表[8.2.1]，但没有帧间预测

561
563

8.2　一些基本的压缩方法

本节描述今天所用的几种主要有损和无误差压缩方法，主要介绍那些主流二值图像、连续色调静止图像和视频压缩标准中业已证实的有用方法。标准本身用于对现有方法进行示范。

8.2.1　霍夫曼编码

消除编码冗余的一种最常用技术要归功于霍夫曼(Huffman [1952])。单独对信源的符号进行编码时，霍夫曼编码对每个信源符号产生最小数量的编码符号。根据香农第一定理(见 8.1.4 节)，对一个固定值 n 生成的编码是最佳的，但限制是每次只能对一个信源符号进行编码。在实践中，信源符号不是图像的灰度，就是一个灰度映射操作的输出(如像素差值、行程等)。

> 参考表 8.3 和表 8.4，霍夫曼编码用于如下压缩标准和其他压缩标准中：
> - CCITT
> - JBIG2
> - JPEG
> - MPEG-1, 2, 4
> - H.261, H.262, H.263, H.264

霍夫曼方法的第一步是，对所考虑符号的概率进行排序，并把具有最小概率的多个符号合并为一个符号，以替代下次信源化简过程中的符号，进而创建一个简化的信源系列。图 8.7 针对二进制编码说明了这种处理(也可以构造 K 元霍夫曼编码)。左侧从上到下列出了一组虚构的信源符号集合及它们的概率（从大到小）。为形成第一次信源简化，底部的两个概率 0.06 和 0.04 合并，形成了一个概率值为 0.1 的“复合符号”。这个复合符号及其对应的概率被置于第一个信源简化列中，以便简化后信源的概率仍按其值从最大到最小排列。这个过程一直重复持续到信源只有两个符号的简化信源(在最右边)为止。

原始信源		信源化简			
符号	概率	1	2	3	4
a_2	0.4	0.4	0.4	0.4	0.6
a_6	0.3	0.3	0.3	0.3	0.4
a_1	0.1	0.1	0.2	0.3	
a_4	0.1	0.1	0.1		
a_3	0.06	0.1			
a_5	0.04				

图 8.7　霍夫曼信源化简

霍夫曼编码过程的第二步是对每个化简后的信源进行编码，从最小的信源开始，直到遍历原始信源。当然，两个符号信源的最小长度的二值码是符号 0 和 1。如图 8.8 所示，这些符号被分配给最右边的两个符号(这种分配是任意的；颠倒 0 和 1 的顺序同样可行)。将简化信源中的两个符号合并到其左侧，生成了概率为 0.6 的简化信源符号，对该符号编码所用的 0 现在分配给了这两个符号，而 0 和 1 是任意加到每个符号上的，以便对它们进行区分。然后，对每个简化信源重复该操作，直到到达原始信源。图 8.8 最左边显示了最终的编码。这个编码的平均长度为

$$L_{\text{avg}} = (0.4)(1) + (0.3)(2) + (0.1)(3) + (0.1)(4) + (0.06)(5) + (0.04)(5) = 2.2 \text{ 比特/像素}$$

该信源的熵为 2.14 比特/符号。

霍夫曼编码过程对一组符号产生最佳编码，其概率服从一次只能对一个符号进行编码的限制。在编码建立之后，编码和/或无误差解码就简单地以查找表的方式完成。编码本身是一种瞬时的、唯一可解码的块编码。之所以称它为块编码，是因为每个信源符号都映射到了一个编码符号的固定序列中。它是瞬时的，因为编码符号串中的每个码字无须参考后续符号就可以进行解码。它是唯一可解码的，因为任何编码符号串只能以一种方式进行解码。因此，任何霍夫曼编码符号串都可以通过从左到右的

564~565

方式对该串中的每个符号进行分析来解码。对于图8.8中的二进制编码，对编码串010100111100从左到右的扫描表明，第一个有效码字为01010，它是符号 a_3 的码；下一个有效编码是011，它对应的符号为 a_1。按这种方式继续下去，可得到完全解码后的消息为 $a_3a_1a_2a_2a_6$。

原始信源			信源化简							
符号	概率	编码	1		2		3		4	
a_2	0.4	1	0.4	1	0.4	1	0.4	1	0.6	0
a_6	0.3	00	0.3	00	0.3	00	0.3	00	0.4	1
a_1	0.1	011	0.1	011	0.2	010	0.3	01		
a_4	0.1	0100	0.1	0100	0.1	011				
a_3	0.06	01010	0.1	0101						
a_5	0.04	01011								

图8.8 霍夫曼编码分配过程

例8.4 霍夫曼编码。

图8.9中的512×512×8比特单色图像具有如图8.9(b)所示的直方图。因为灰度不是等概率的，霍夫曼编码过程的MATLAB实现用7.428比特/像素对其编码——包括重建原始8比特图像灰度所要求的霍夫曼编码表。由 $512^2\times(7.428-7.3838)$ 或11 587比特可知，压缩后的表示超过了图像的估计熵［由式(8.1-7)得出估计熵为7.3838比特/像素］约0.6%。得到的压缩率和对应的相对冗余度分别是 $C = 8/7.428 = 1.077$ 和 $R = 1-(1/1.077) = 0.0715$。这样，原始8比特定长灰度表示的7.15%就作为编码冗余而被去除。

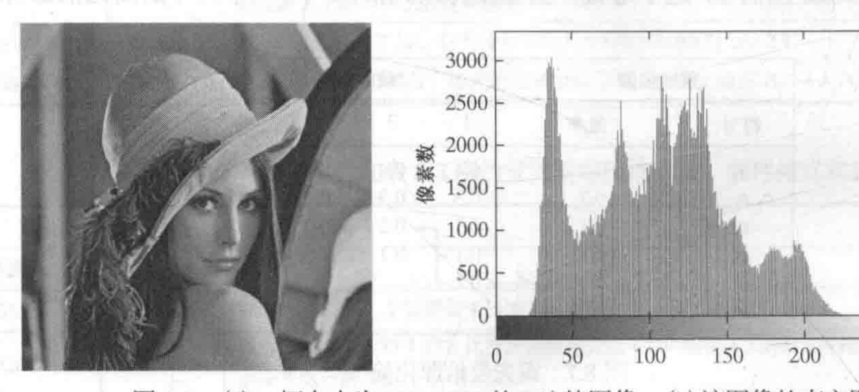


图8.9 (a)一幅大小为512×512的8比特图像；(b)该图像的直方图

对大量符号编码时，最佳霍夫曼编码的构造并不简单。对于有 J 个信源符号的通常情况，需要 J 个符号概率、$J-2$ 次信源简化和 $J-2$ 次编码赋值。事先可以估计信源符号的概率时，使用预计算的霍夫曼编码可以实现“接近最佳”编码。一些通用的图像压缩标准，包括8.2.8节和8.2.9节中讨论的JPEG和MPEG标准，都规定了默认的霍夫曼编码表，这些编码表是以实验数据为基础预先计算出来的。

8.2.2 Golomb 编码

本节介绍具有指数衰减概率分布输入的非负整数编码。通过使用计算上比霍夫曼编码还要简单的一类编码，这种类型的输入可被最佳化编码(在香农第一定理的意义上)。编码本身第一次是针对非负行程的表示提出来的(Golomb[1966])。在接下来的讨论中，符号 $\lfloor x \rfloor$ 表示小于等于 x 的最大整数，$\lceil x \rceil$ 表示大于等于 x 的最小整数，$x \bmod y$ 表示 x 被 y 除的余数。

参考表8.3和表8.4，Golomb编码用于JPEG-LS压缩和AVS压缩。

565~566

给定一个非负整数 n 和一个正整数除数 $m>0$ 后，表示为 $G_m(n)$ 的 n 关于 m 的 Golomb 编码是商 $\lfloor n/m \rfloor$ 的一元编码和 $n \bmod m$ 的二进制表示的并集。$G_m(n)$ 的构建如下：

步骤 1　形成商 $\lfloor n/m \rfloor$ 的一元编码(整数 q 的一元编码定义为 q 个 1 紧跟着一个 0)。

步骤 2　令 $k=\lceil \log_2 m \rceil, c=2^k-m, r=n \bmod m$，并计算截短的余数 r'，例如，使其满足

$$r'=\begin{cases} r\text{截短至 } k\text{–}1 \text{ 比特}, & 0 \leqslant r < c \\ r+c\text{ 截短至 } k \text{ 比特}, & \text{其他} \end{cases} \tag{8.2-1}$$

步骤 3　连接步骤 1 和步骤 2 的结果。

例如，为计算 $G_4(9)$，可从求商 $\lfloor 9/4 \rfloor = \lfloor 2.25 \rfloor = 2$ 的一元编码开始，它的一元编码是 110(步骤 1 的结果)。然后，令 $k=\lceil \log_2 4 \rceil = 2, c=2^2-4=0, r=9 \bmod 4$，在二进制中，是 1001 mod 0100 或 0001。根据式(8.2-1)，r' 是 r(即 0001)截短到 2 比特的结果，它是 01(步骤 2 的结果)。最后，连接来自步骤 1 的 110 和来自步骤 2 的 01，得到 11001，即 $G_4(9)$。

对于 $m=2^k$，$c=0$ 的特殊情况，对所有 n，在式(8.2-1)中，$r'=r=n \bmod m$ 截短至 k 比特。产生 Golomb 编码所要求的除法变成二进制移位操作和计算上更简单的编码，这种编码称为 Golomb-Rice 码或 Rice 码(Rice[1975])。表 8.5 中的 2、3 和 4 列列出了前 10 个非负整数的 G_1, G_2 和 G_4 码。因为在每种情况下 m 都是 2 的幂(即 $1=2^0, 2=2^1$ 和 $4=2^2$)，所以它们也是前三个 Golomb-Rice 码。此外，G_1 是非负整数的一元码，因为对所有 n，有 $\lfloor n/1 \rfloor = n$ 和 $n \bmod 1 = 0$。

记住，Golomb 码只能用于表示非负整数，并且有许多 Golomb 码可供选择，其有效应用中的一个关键步骤是除数 m 的选择。当被表示的整数具有概率质量函数(PMF)[①]的几何分布时

$$P(n)=(1-\rho)\rho^n \tag{8.2-2}$$

> 由式(8.2-2)中的 PMF 定义的离散概率分布称为几何概率分布，其对应的连续形式是指数分布。

式中 $0<\rho<1$，可以证明 Golomb 码是最佳的，在该意义上，即当

$$m=\left\lceil \frac{\log_2(1+\rho)}{\log_2(1/\rho)} \right\rceil \tag{8.2-3}$$

时，$G_m(n)$ 为所有唯一可判读的编码提供了最短的平均码长(Gallager and Voorhis[1975])。

> PMF 的图形表示是直方图。

图 8.10 画出了式(8.2-2)在三个不同 ρ 值时的曲线，并以图形方式说明了 Golomb 编码处理得较好的符号概率(即编码效率)。如图所示的那样，小整数的可能性要比大整数大得多。

因为一幅图像中的灰度的概率［如图8.9(b)的直方图］未必与式(8.2-2)中规定的概率和图8.10(a)中所示的概率匹配，所以 Golomb 码很少用于灰度编码。但对灰度差编码时，所得“差值”的概率(见 8.2.9 节)——值得注意的是负差值这一例外——常常类似于式(8.2-2)和图 8.10(a)。为了在仅能表示非负整数的 Golomb 编码中处理负差值，通常使用一个类似于下式的映射：

$$M(n)=\begin{cases} 2n, & n \geqslant 0 \\ 2|n|-1, & n<0 \end{cases} \tag{8.2-4}$$

例如，使用该映射，图 8.10(b)中所示的双边 PMF 可变换为图 8.10(c)中的单边 PMF。整数被重新排序，即交换负整数和正整数，所以负整数可被映射到奇正整数位置。如果 $P(n)$是双边的且中心位

① 概率质量函数(PMF)定义一个离散随机变量完全等于某个值的概率。PMF 与 PDF(概率密度函数)的不同之处在于，PDF 的值不是概率；相反，PDF 在指定区间上的积分才是概率。

566 ~ 568

于零处，那么$P(M(n))$将是单边的。然后，被映射的整数$M(n)$可以使用一种合适的 Golomb-Rice 码来有效地编码(Weinberger et al.[1996])。

表 8.5 整数 0~9 的几种 Golomb 码

N	$G_1(n)$	$G_2(n)$	$G_4(n)$	$G_{\exp}^0(n)$
0	0	00	000	0
1	10	01	001	100
2	110	100	010	101
3	1110	101	011	11000
4	11110	1100	1000	11001
5	111110	1101	1001	11010
6	1111110	11100	1010	11011
7	11111110	11101	1011	1110000
8	111111110	111100	11000	1110001
9	1111111110	111101	11001	1110010

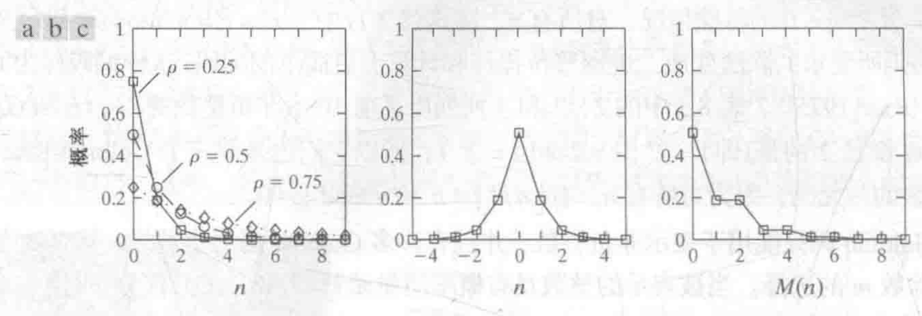


图 8.10 (a)来自式(8.2-2)的三个单边的几何分布；(b)一个双边的指数衰减分布；(c)使用式(8.2-4)重新对图(b)重排序的结果

例 8.5 Golomb-Rice 编码。

再次考虑来自图 8.1(c)的图像，并注意到其直方图［见图 8.3(a)］类似于图 8.10(b)的双侧分布。如果我们令 n 是图像中的非负整数灰度，$0 \leqslant n \leqslant 255$，$\mu$ 是平均灰度，$P(n-\mu)$ 是如图 8.11(a)所示的双侧分布。这条曲线是用图像中的像素总数来归一化图8.3(a)中的直方图，并将归一化后的值向左移动 128(效果上相当于从图像中减去平均灰度)产生的。根据式(8.2-4)，$P(M(n-\mu))$ 是图8.11(b)所示的单侧分布。如果使用表 8.5 中第二列的 G_1码的 MATLAB 实现对重新排序后的灰度值进行 Golomb 编码，则编码后的表示比原图像小 4.5 倍(即 $C=4.5$)。使用变长编码，G_1码在理论上实现 4.5/5.1 或 88%的压缩率是可能的(基于例 8.2 中计算的熵，变长编码可实现的最大压缩率是 $C=8/1.566 \approx 5.1$)。此外，Golomb 编码可达到用霍夫曼编码 MATLAB 实现的压缩的 96%，并且不需要定制霍夫曼编码表的计算。

现在考虑图 8.9(a)中的图像。如果它的灰度已用上面谈到的相同 G_1码进行了 Golomb 编码，那么 $C=0.0922$，即存在数据扩展。这是由于图 8.9(a)中图像的灰度的概率与式(8.2-2)定义的概率相比，有很大的不同。以相同的方法，当使用概率与由计算得到的编码的概率不同的编码符号时，霍夫曼编码可能产生数据扩展。实践中我们可以不按照输入概率的假定来设计编码，但压缩性能变差和数据扩展的风险会更大。

> 在式(8.1-2)中，当 C 小于 1 时，就存在数据扩展。

为结束对 Golomb 编码的讨论，我们注意到，表 8. 5 第 5 列包含的表示为 $G_{\exp}^0(n)$ 的第 0 阶指数 Golomb 码的前 10 个码。指数 Golomb 编码对于行程编码很有用，因为短行程和长行程都可以有效地编码。k 阶指数 Golomb 编码 $G_{\exp}^k(n)$ 计算如下：

568～569

步骤 1　寻找一个整数 $i, i \geqslant 0$，使其满足

$$\sum_{j=0}^{i-1} 2^{j+k} \leqslant n < \sum_{j=0}^{i} 2^{j+k} \tag{8.2-5}$$

并形成 i 的一元码。如果 $k=0$，则 $i = \lfloor \log_2(n+1) \rfloor$，该编码也是周知的 Elias gamma 编码。

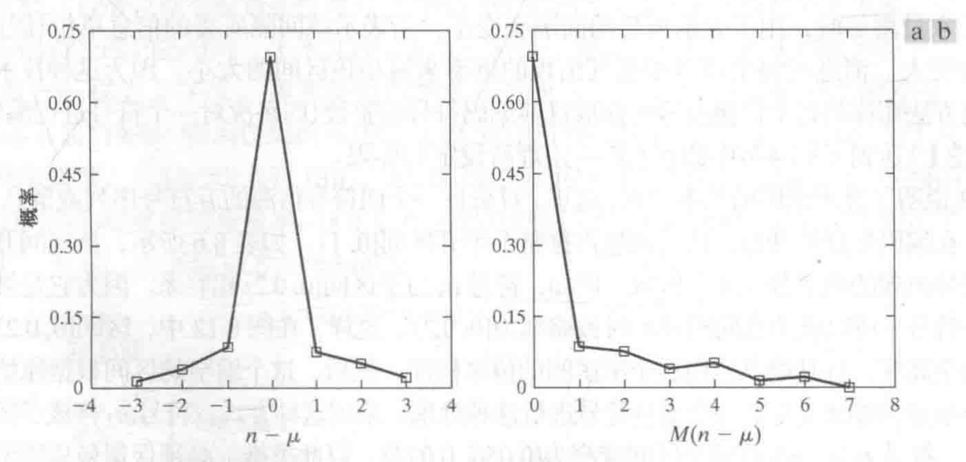


图 8.11　(a) 从图 8.1(c) 中图像的每个像素中减去平均灰度后该图像的概率分布；(b) 对图 (a) 使用式 (8.2-4) 得到的映射

步骤 2　把二进制表示

$$n - \sum_{j=0}^{i-1} 2^{j+k} \tag{8.2-6}$$

截短到 $k+i$ 个最低阶比特。

步骤 3　连接步骤 1 和步骤 2 的结果。

例如，为求 $G_{\exp}^0(8)$，我们在步骤 1 中令 $i = \lfloor \log_2 9 \rfloor$ 或 3，因为 $k=0$。然后，式 (8.2-5) 得到满足，因为

$$\sum_{j=0}^{3-1} 2^{j+0} \leqslant 8 < \sum_{j=0}^{3} 2^{j+0}$$

$$\sum_{j=0}^{2} 2^{j} \leqslant 8 < \sum_{j=0}^{3} 2^{j}$$

$$2^0 + 2^1 + 2^2 \leqslant 8 < 2^0 + 2^1 + 2^2 + 2^3$$

$$7 \leqslant 8 < 15$$

3 的一元码是 1110，步骤 2 中的式 (8.2-6) 产生

$$8 - \sum_{j=0}^{3-1} 2^{j+0} = 8 - \sum_{j=0}^{2} 2^{j} = 8 - (2^0 + 2^1 + 2^2) = 8 - 7 = 1 = 0001$$

当截短至其 3 + 0 个最低阶比特时，它变为 001。连接步骤 1 和步骤 2 的结果得到 1110001。注意，这是 $n = 8$ 时表 8.5 中第 4 列的条目。最后，我们注意到，类似于上一节的霍夫曼编码，表 8.5 中的 Golomb 编码是变长的、瞬时唯一可解码的块编码。

569 ~ 570

8.2.3 算术编码

与前两节的变长编码不同，算术编码生成的是非块码。算术编码可以追溯至 Elias 的研究工作(见 Abramson[1963])，信源符号和码字之间不存在一一对应的关系。相反，算术编码给信源符号(或消息)的整个序列分配了单一的算术码字。这个码字本身定义了一个介于 0 和 1 之间的实数间隔。当消息中的符号数量增加时，用于表示消息的间隔会变小，而表示该间隔所需的信息单位(假设为比特)的数量则会变大。消息的每个符号根据其出现的概率来减小该区间的大小。因为这种技术不要求像霍夫曼编码方法那样将每个信源符号转换成这些编码符号的整数(即每次对一个符号进行编码)，所以它(仅在理论上)达到了8.1.4节中的香农第一定理所设定的界限。

图 8.12 说明了算术编码的基本过程。这里，对来自一个四符号信源的五符号序列或消息 $a_1a_2a_3a_3a_4$ 进行编码。在编码处理的开始，假设消息占据整个半开区间[0, 1)。如表 8.6 所示，该区间开始时根据每个信源符号出现的概率分为 4 个区域。例如，符号 a_1 与子区间[0, 0.2)相联系。因为它是被编码的消息的第一个符号，所以该消息间隔开始时被缩窄为[0, 0.2)。这样，在图 8.12 中，区间[0, 0.2)就被扩展到该图形的全高度，且其端点使用这个窄区间的值来标注。然后，这个缩窄的区间根据原始信源符号的概率进行细分，并继续对下一个消息符号进行这种处理。采用这种方式，符号 a_2 将该子区间变窄为[0.04, 0.08)，符号 a_3 进一步将该子区间变窄为[0.056, 0.072)，以此类推。必须保留最后的消息符号，以作为特定的消息结束指示符，它将子区间变窄为[0.067 52, 0.068 8)。当然，在这个子区间内的任何数字(如 0.068)都可以用来表示该消息。

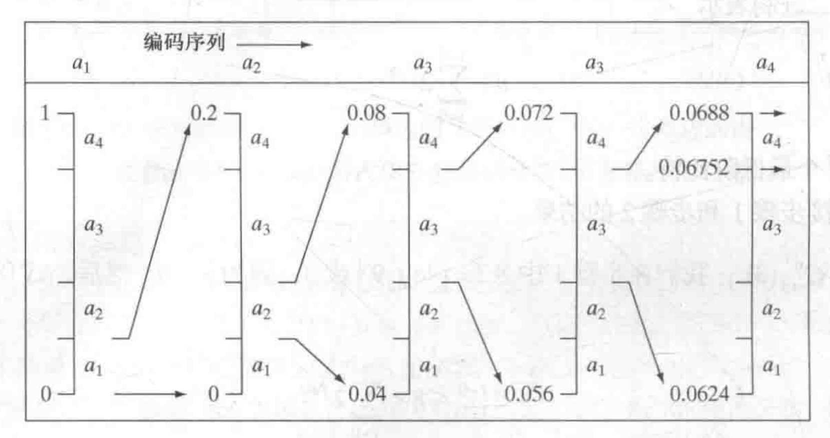


图 8.12 算术编码过程

在图 8.12 所示算术编码的消息中，使用了 3 个十进制数字来表示这个五符号消息。这就转换为每信源符号 0.6 个十进制数字，而由式(8.1-6)，与之相比的信源的熵为每个信源符号 0.58 个十进制数字。当被编码的序列的长度增加时，得到的算术编码接近香农第一定理所设定的界限。实际上，有两个因素会使得编码性能无法达到这一界限：(1)为了将一个消息与其他消息分开，增加了消息结束指示符；(2)所用算法的精度是有限的。算术编码的实际实现通过引入一种缩放策略和一种舍入策略解决第二个问题(Langdon and Rissanen[1981])。缩放策略是在根据符号出现的概率细分每个子区间之前，把每个子区间重新归一化到区间[0, 1)。舍入策略则保证根据算法的有限精度进行截短，这不会影响被编码子区间的准确表示。

表 8.6 算术编码示例

信源符号	概率	初始子区间
a_1	0.2	[0.0, 0.2)
a_2	0.2	[0.2, 0.4)
a_3	0.4	[0.4, 0.8)
a_4	0.2	[0.8, 1.0)

570 ~ 571

自适应上下文相关的概率估计

使用精确的输入符号概率模型，即提供被编码符号的真实概率的模型，算术编码器在要求表示被编码符号平均码字符号数最小的情形下是最佳的。然而，像霍夫曼编码和 Golomb 编码那样，不精确的概率模型可导致非最佳结果。改进所采用概率精确性的一种简单方法是，使用自适应的、上下文相关的概率模型。自适应概率模型会像符号已被编码或变得已知那样来更新符号的概率。这样，概率就会适应被编码符号的局部统计。上下文相关的模型提供了概率，该概率基于被编码符号周围的预定义的邻域像素——称为上下文。通常使用因果关系的上下文，即限制已被编码的符号。Q 编码器 (Pennebaker et al.[1988]) 和 MQ 编码器 (ISO/IEC[2000]) 两种周知的算术编码技术使用了自适应和上下文相关的概率模型，它们已引入到 JBIG、JPEG-2000 和其他重要图像压缩标准中。Q 编码器在区间重新归一化期间动态地更新符号概率，这是算术编码处理的一部分。Golomb 编码中也使用了自适应上下文相关模型，例如，在 JPEG-LS 压缩标准中。

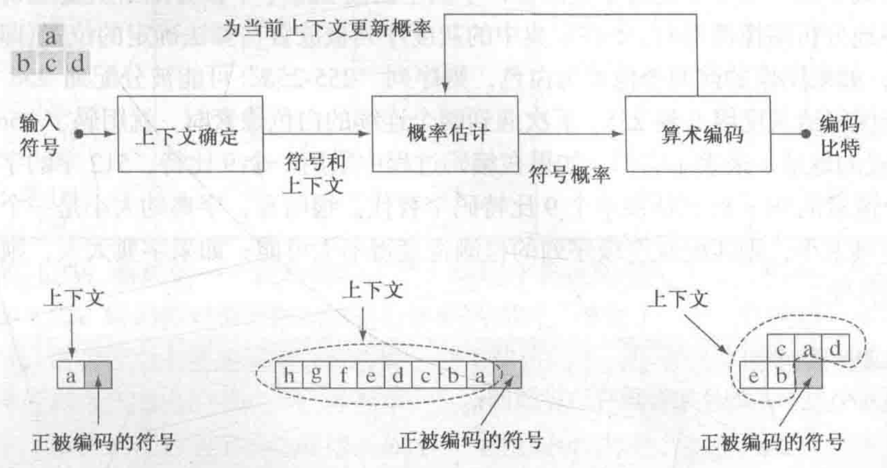


图 8.13　(a) 自适应上下文相关的算术编码方法（通常用于二值信源符号）；(b)～(d) 三种可能的上下文模型

图 8.13(a) 给出了二值信源符号的自适应上下文相关算术编码包括的步骤。当对二值符号编码时，通常使用算术编码。当对每个符号（或比特）开始编码过程时，其上下文由图 8.13(a) 中的上下文决定模块形成。图 8.13(b) 到 (d) 显示了 3 种可被使用的上下文：(1) 前一符号；(2) 前一组符号；(3) 前一些符号加上前一扫描行上的符号。对所示的三种情况，概率估计模块必须管理 2^1（或 2）个、2^8（或 256）个和 2^5（或 32）个上下文及与它们相关联的概率。例如，如果使用了图 8.13(b) 中的上下文，就必须跟踪传统的概率 $P(0|a=0)$（被编码符号为 0 的概率给出了前一符号为 0）、$P(1|a=0)$、$P(0|a=1)$ 和 $P(1|a=1)$。然后，将合适的概率作为当前上下文的函数传到算术编码模块，并根据图 8.12 中说明的过程，产生算术编码输出序列。然后，更新包含在当前编码步骤中的相关概率，以反映该上下文中另一个符号已被处理的事实。

最后，我们注意到许多算术编码技术受美国专利的保护（可能还受其他法律保护）。因为侵害这些专利可能导致高额罚款，包含有霍夫曼编码和算术编码选项的 JPEG 标准的多数实现，通常仅支持霍夫曼编码。

8.2.4　LZW 编码

> 参考表 8.3 和表 8.4，LZW 编码用于 GIF、TIFF 和 PDF 文件格式，但不用于任何国际压缩标准批准的文件格式。

前几节覆盖的技术集中在消除编码冗余上。本节考虑一种无误差压缩图像中的空间冗余的方法。称为 Lempel-Ziv-Welch (LZW) 编码的这种技术，将定长码字分配给变长信源符号序列。回忆 8.1.4 节，在第一定理的证明中，

572～573

香农使用了信源符号编码序列这一思想，而不是单个信源符号。LZW 编码的关键特点是，它不需要被编码符号出现的概率的先验知识。尽管直到现在这种压缩技术还在美国专利的保护之下，但 LZW 压缩技术已被引入许多主流图像文件格式中，包括 GIF、TIFF 和 PDF 等。创建 PNG 格式得到了 LZW 的许可。

例 8.6　图 8.9(a)的 LZW 编码。

再次考虑图 8.9(a)中大小为 512×512 的 8 比特图像。使用 Adobe Photoshop，该图像的未压缩 TIFF 版本要求 286 740 字节的磁盘空间——512×512 的 8 比特像素加上 24 596 字节的头文件。然而，用 TIFF 的 LZW 压缩选项得到的文件大小为 224 420 字节。压缩率是 $C = 1.28$。回忆例 8.4 中图 8.9(a)的霍夫曼编码表示，$C = 1.077$。由 LZW 方法实现的额外压缩是由于消除了图像的某些空间冗余。

LZW 编码在概念上非常简单(Welch[1984])。在编码过程的开始阶段，先构建一个包含被编码信源符号的码书或字典。对于 8 比特单色图像，字典中的前 256 个字被分配给灰度值 0, 1, 2, …, 255。当编码器顺序地分析图像像素时，不在字典中的灰度序列被放置在算法确定的位置(即下一个未用的位置)。例如，如果图像的前两个像素为白色，则序列“255-255”可能被分配到 256 的位置，该位置后面的地址保留给灰度级 0 到 255。下次遇到两个连续的白色像素时，就用码字 256（即包含序列 255-255 的位置的地址）来表示它们。如果在编码过程中采用一个 9 比特、512 字的字典，则最初用于表示这两个像素的(8 + 8)比特被单个 9 比特码字替代。很明显，字典的大小是一个很重要的系统参数。如果字典太小，则匹配灰度级序列的检测将变得不太可能；如果字典太大，则码字的大小反而会影响压缩性能。

例 8.7　LZW 编码。

考虑下列大小为 4×4 的 8 比特垂直边缘图像：

$$\begin{matrix} 39 & 39 & 126 & 126 \\ 39 & 39 & 126 & 126 \\ 39 & 39 & 126 & 126 \\ 39 & 39 & 126 & 126 \end{matrix}$$

表 8.7 中详细给出了对其 16 个像素进行编码的步骤。假设一个 512 字带有下列初始内容的字典：

字典位置	条　目
0	0
1	1
⋮	⋮
255	255
256	—
⋮	⋮
511	—

位置 256 到 511 开始时未使用。

图像按从左到右、从上到下处理像素的方法来进行编码。每个连续的灰度值与表 8.7 中称之为“当前可识别的序列”的第 1 列的一个变量链接。就像我们看到的那样，该变量最初为空。对每个链接的序列搜索字典，如果找到，如表中第一行的情况，则用新链接且可识别的序列(即位于字典中的序列)替代它。这可在第 2 行的第 1 列中实现。这既不产生输出代码，也不更改字典。然而，如果没有找到链接的序列，则将当前可识别序列的地址作为下一个编码值输出，并将链接但不可识别的序列添加到字典中，当前可识别的序列则初始化为当前的像素值。这出现在表中的第 2 行。最后两列详细给出了扫描整个 4×4

573~574

图像时添加到了字典中的灰度序列。定义了 9 个附加的码字。编码结束时，字典包含 256 个码字，且 LZW 算法成功地识别了几个重复的灰度序列——平衡它们将 128 比特的原图像降到了 90 比特(即 10 个 9 比特码)。从上到下读取第 3 列即可得到编码输出。得到的压缩率是 1.42:1。

表 8.7　LZW 编码示例

当前可识别的序列	正被处理的像素	编码后的输出	字典位置(码字)	字 典 词 条
	39			
39	39	39	256	39-39
39	126	39	257	39-126
126	126	126	258	126-126
126	39	126	259	126-39
39	39			
39-39	126	256	260	39-39-126
126	126			
126-126	39	258	261	126-126-39
39	39			
39-39	126			
39-39-126	126	260	262	39-39-126-126
126	39			
126-39	39	259	263	126-39-39
39	126			
39-126	126	257	264	39-126-126
126		126		

刚刚示范的 LZW 编码的一个独特的特点是，编码字典或码书是在对数据进行编码的同时创建的。很明显，在 LZW 解码器对编码数据流进行解码的同时，建立了一个同样的解压缩字典。作为练习(见习题 8.20)，请读者对上例的输出进行解码，并重建该码书。尽管在该例中不需要，但多数实际应用都需要一种处理字典溢出的策略。一种简单的解决办法是，当字典已满时，刷新或重新初始化该字典，并使用一个新初始化后的字典继续编码。另一种复杂的选择是，监控压缩性能，并在性能变得低下或不可接受时刷新字典。此外，需要时，可跟踪并替换那些使用最少的字典词条。

8.2.5　行程编码

像 8.1.2 节中解释的那样，沿其行(或列)重复灰度的图像，通常可用相同灰度的行程表示为行程对来压缩，其中每个行程对指定一个新灰度的开始和具有该灰度的连续像素的数量。这种称为行程编码的技术，是 20 世纪 50 年代发展起来的，目前连同其二维扩展一起，已成为传真编码标准的压缩方法。压缩是通过消除空间冗余的一种简单形式(即一组相同的灰度)来实现的。当相同像素的行程较小(或没有)时，行程编码会导致数据扩展。

参考表 8.3 和表 8.4，行程编码用于 CCITT、JBIG2、JPEG、M-JPEG、MPEG-1, 2, 4、BMP 和其他压缩标准与文件格式中。

例 8.8　BMP 文件格式中的 RLE。

BMP 文件格式使用行程编码的一种形式，其中图像数据以两种不同的模式来表示：编码和绝对值，任何一种模式在图像的任何地方都会出现。在编码模式中，用两个字节 RLE 表示。第一个字节规定了连续像素的数目，它有包含在第二个字节中的彩色索引。8 比特彩色索引从 256 种可能的灰度表中选择该行程的灰度(彩色或灰度值)。

在绝对值模式中，第一个字节是 0，第二个字节通知 4 个可能的条件之一，如表 8.8 所示。当第二个字节是 0 或 1 时，就到达了一行或一幅图像的末尾。如果第二个字节是 2，则下两个字节包含相对图像中一个新空间位置的无符号水平偏移和垂直偏移。如果第二个字节在 3 和 255 之间，则它指定了非压缩像素的数量——紧跟着包含有一个像素的彩色索引的每个后续的字节。总字节数必须与一个 16 比特字的边界来排列。

574
~
576

图 8.9(a)中所示的 512×512×8 比特图像的未压缩 BMP 文件(使用 Photoshop 存储)要求 263 244 字节的存储空间。使用 BMP 的 RLE 选项压缩后，文件扩展为 267 706 字节，压缩率 C 为 0.98，没有足够的使得行程压缩有效的相等灰度的行程，导致了少量的扩展。然而，对于图 8.1(c)中的图像，BMP 的 RLE 选项得到的压缩率是 $C = 1.35$。

注意，由于开销的不同，在例 8.7 中，未压缩的 BMP 文件小于未压缩的 TIFF 文件。

压缩二值图像时，行程编码特别有效。因为仅有两种可能的灰度(黑和白)，所以邻近像素的灰度更可能是相同的。另外，每一图像行可仅用长度序列来表示，而不是例 8.8 中所用的长度、灰度对。其基本概念是，对从左到右扫描一行时所遇到的 1 或 0 的每个连续组，使用这些连续组的长度进行编码，并且建立决定行程值的约定。最通常的约定是：(1) 指定每行的第一个行程的值，或(2) 假设每行从一个白色行程开始，其行程事实上可能是零。

表 8.8　BMP 绝对值编码模式选项。在这种模式中，BMP 对的第一个字节为 0

第二个字节值	条　件
0	行尾
1	图像尾
2	移到一个新位置
3~255	逐个指定像素

尽管行程编码本身是一种压缩二值图像的有效方法，但对行程本身进行变长编码可以实现额外的压缩。黑和白行程可个别地进行编码，这些变长编码是根据它们自身的特殊统计数据制定的。例如，令符号 a_j 表示一个长度为 j 的黑色行程，我们可以估计由一个假想的黑色行程长度信源发出的符号 a_j 的概率，它等于整个图像中长度为 j 的黑色行程数除以黑色行程总数。这个黑色行程信源的熵的估计表示为 H_0，然后，将这些概率代入式(8.1-6)。对于白色行程的熵，可得出类似的结论，表示为 H_1。该图像的近似行程熵为

$$H_{RL} = \frac{H_0 + H_1}{L_0 + L_1} \tag{8.2-7}$$

式中，变量 L_0 和 L_1 分别表示黑色行程和白色行程的平均值。式(8.2.7)提供了对二值图像的行程使用变长编码时所要求的每像素平均比特数的估计。

两种最古老且应用最广的二值图像压缩标准是 CCITT Group 3 和 4。虽然它们已被用于各种计算机应用中，但它们最初还是针对电话网传输文件而设计的传真编码方法。Group 3 标准使用一维行程编码技术，其中，每组 K 行($K = 2$ 或 4)中的最后 K-1 行可随意地以一种二维编码方法来编码。Group 4 标准是 Group 3 标准的一种简化或改进版本，该标准中只允许进行二维编码。两个标准都使用同一种二维编码方法，从用前一行的信息对当前行进行编码的意义上看，这种方法是二维的。下面讨论一维编码和二维编码。

一维 CCITT 压缩

在一维 CCITT Group 3 压缩标准中，一幅图像的每一行①被编码为一系列变长霍夫曼码字，在从左到右对该行的扫描中，这些码字表示交替的白色行程和黑色行程。采用的压缩方法通常称为改进的霍夫曼(MH)编码方法。码字本身有两种类型，在该标准中分别称为终结码和补偿码。如果行程 r 小于 63，则使用来自附录 A 的表 A.1 中的终结码来表示它。注意，对于白色行程和黑色行程，该标准规定了不同的终结码。如果 $r > 63$，则使用两种码——对商 $\lfloor r/64 \rfloor$ 使用补偿码，而对 r mod 64 的余数使用终结码。表 A.2 中列出了补偿码，并且，这些补偿码可能依赖，也可能不依赖于被编码的行程的灰度(黑或白)。如果 $\lfloor r/64 \rfloor < 1792$，则规定分隔黑色行程和白色行程的补偿码；否则，补偿码是独立于行程灰

回忆 8.2.2 节可知，符号 $\lfloor x \rfloor$ 表示小于等于 x 的最大整数。

576~577

① 在该标准中，图像称为页面，图像序列则称为文档。

度的。标准要求每一行以白色行程码字开始，事实上它可能是 00110101，即长度为零的白色行程的码字。最后，用一个唯一的行的结尾(EOL)码字 000000000001 来终止每一行，并对每幅新图像的第一行发出信号。图像序列的结尾由 6 个连续的 EOL 来指出。

二维 CCITT 压缩

CCITT Group 3 标准和 CCITT Group 4 标准采用的二维压缩方法是逐行方法，在这种方法中，对每个黑到白或白到黑的行程转换的位置，根据位于当前编码行上的参考元素 a_0 的位置来编码。已编码的前一行称为参考行；每幅新图像的第一行的参考行是一个虚构的白色行。所用的二维编码技术称为相对元素地址指派(READ)编码。在 Group 3 标准中，相继的 MH 编码行间允许 1 个或 3 个 READ 编码行，并且这种技术称为改进的 READ(MR)编码。在 Group 4 标准中，允许较大数量的 READ 编码行，这种方法称为改进的改进的 REDA(MMR)编码。如前面解释的那样，从使用来自前一行的信息对当前行进行编码的意义上看，这种编码是二维的。不涉及二维变换。

图8.14说明了对单一扫描行的基本二维编码过程。注意，该过程的初始几步指向几个关键的变化元素的定位点处：a_0, a_1, a_2, b_1 和 b_2。标准把变化元素定义为一个像素，该像素的值不同于同一行上前一个像素的值。最重要的变化元素是 a_0(参考元素)，它不是每个新编码行的第一个像素左侧的虚构白色变化元素所置的位置，就是由前一种编码模式决定。编码模式在下一段讨论。a_0 被定位后，a_1 就可标识为当前编码行上 a_0 右侧的下一个变化元素的位置，a_2 可标识为编码行上 a_1 右侧的下一个变化元素的位置，b_1 可标识为与 a_0 具有相反值的变化元素，并在参考行(或前一行)上 a_0 的右侧，b_2 作为下一个变化元素可标识为参考行上 b_1 的右侧。如果这些变化元素中的任意一个都未被检测到，则把它们放到一个合适的行上最后一个像素右侧的一个虚构像素的位置。图 8.15 提供了各个变化元素间的通常关系的两个说明。

当前参考元素和相关的变化元素标识出来后，执行两个简单的测试，以便选择三种可能的编码模式中的一种：通过模式、垂直模式和水平模式。初始测试对应于图 8.14 流程图的第一个分支点，比较 b_2 和 a_1 的位置。第二个测试，对应于图 8.14 中的第二个分支点，计算 a_1 的位置和 b_1 的位置之间的距离(以像素计算)，并与 3 进行比较。然后，依照这些测试的结果，将测试结果输入到图8.14中的对应编码框图中，并执行合适的编码过程。这样，对于每个流程图，就建立了一个新的参考元素，为下一次编码反复做好准备。

表8.9对三种可能的编码模式定义了指定用途的编码。在通过模式中，特别排除了 b_2 直接在 a_1 之上的情况，仅需要通过模式码字0001。如图8.15(a)所示，该模式可识别与当前白或黑编码行行程不重叠的白或黑参考行行程。在水平编码模式中，从 a_0 到 a_1 的距离和从 a_1 到 a_2 的距离必须根据附录 A 中的表 A.1 和表 A.2 的终结码和补偿码来编码，然后，添加到水平模式码字 001 上。这在表 8.9 中由符号 $001 + M(a_0a_1) + M(a_1a_2)$ 指出，其中 a_0a_1 和 a_1a_2 分别表示从 a_0 到 a_1 的距离和从 a_1 到 a_2 的距离。最后，在垂直编码模式中，6 个特殊变长码之一分配给了 a_1 和 b_1 之间的距离。图 8.15(b)说明了水平模式和垂直模式编码中包含的参数。表 8.9 底部的扩展模式码字用于加入一种可选的传真编码模式。例如，0000001111 码用于启动一种传输的非压缩模式。

例 8.9　CCITT 垂直模式编码实例。

虽然图 8.15(b)对水平和垂直模式编码的参数做了注释(有助于上面的讨论)，但描述的黑像素和白像素模式是垂直模式编码的一种情况。也就是说，因为 b_2 在 a_1 的右侧，图 8.14 中的第一个(或通过模式)测试失败了，确定是引入垂直模式还是水平模式的第二次测试指出应使用垂直模式编码，因为从 a_1 到 b_1 的距离小于 3。根据表 8.9，合适的码字是 000010，这意味着 a_1 是 b_1 左侧的两个像素。在下一次编码反复的准备中，a_0 被移到 a_1 的位置。

577 ~ 581

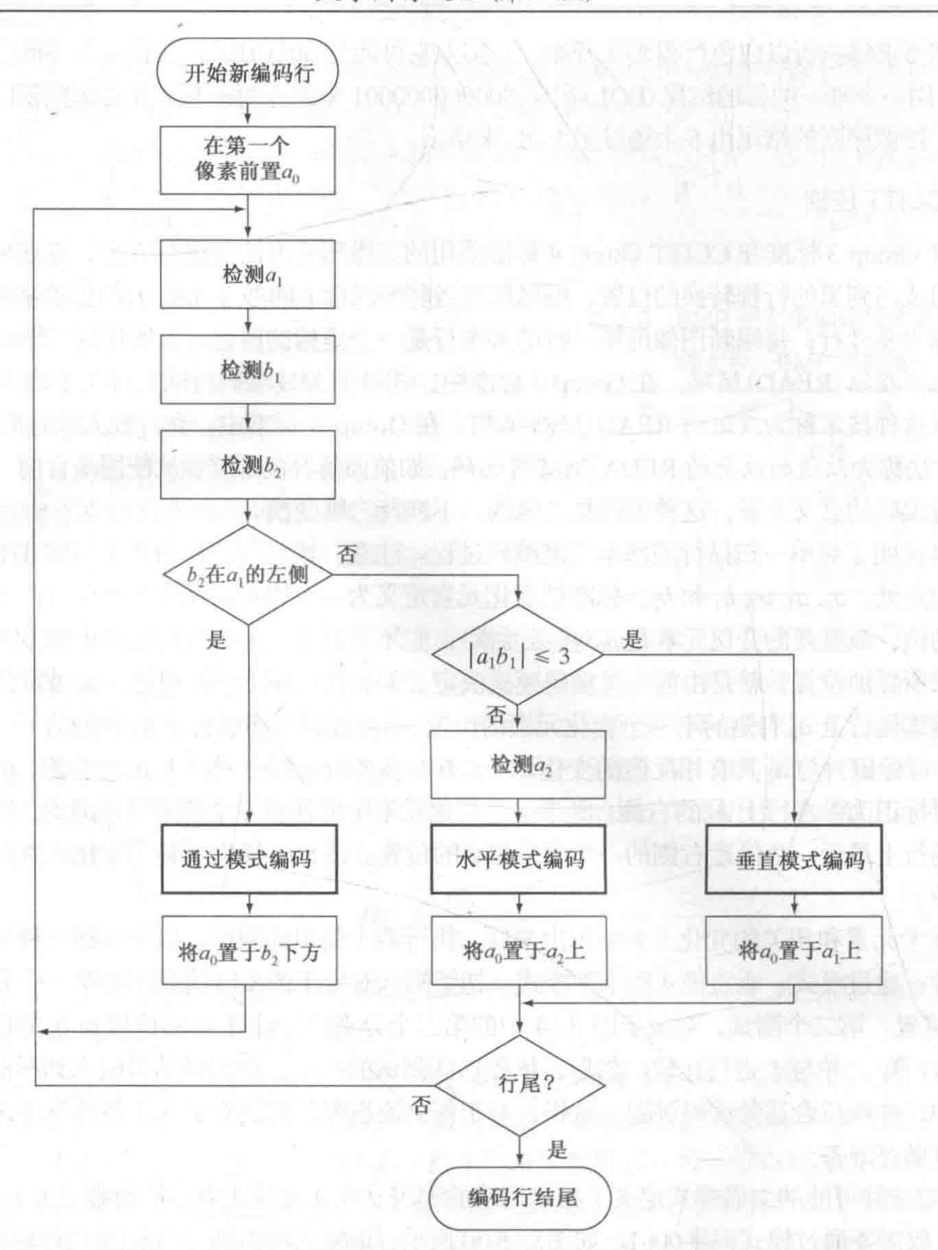


图 8.14 CCITT 二维 READ 编码过程。符号$|a_1b_1|$表示变化元素 a_1 和 b_1 间距离的绝对值

表 8.9 CCITT 二维编码表

模 式	码 字	模 式	码 字
通过	0001	a_1 是 b_1 右侧的 3 个像素	0000011
水平	$001 + M(a_0a_1) + M(a_1a_2)$	a_1 是 b_1 左侧的 1 个像素	010
垂直		a_1 是 b_1 左侧的 2 个像素	000010
a_1 在 b_1 下方	1	a_1 是 b_1 左侧的 3 个像素	0000010
a_1 是 b_1 右侧的 1 个像素	011	扩展	0000001xxx
a_1 是 b_1 右侧的 2 个像素	000011		

例 8.10 CCITT 压缩实例。

580 ~ 581

图 8.16(a)是以大约 1/3 大小显示的 7×9.25 英寸书页的 300 dpi 扫描图像。注意，该页约有一半包含正文，半色调图像约占 9%，其余为空白。图 8.16(b)是被放大的书页的一部分。记住，我们处理的是二

值图像；如4.5.4 节描述的那样，在打印时使用半色调处理方法创建了灰色调的幻影。如果图 8.16(a)中图像的二值像素以每字节 8 像素组的形式存储，则通常称为文档的 1952×2697 比特扫描图像需要 658 068 字节。该文档的非压缩 PDF 文件(由 Photoshop 创建)需要 663 445 字节。CCITT Group 3 压缩可将该文件减少到 123 497 字节，压缩率为 $C = 5.37$；CCITT Group 4 压缩将该文件减少到 110 456 字节，压缩率约增加到 6。

> 不要将此处使用的 PDF(便携文档格式)与前几节及前几章中表示概率密度函数的 PDF 混淆。

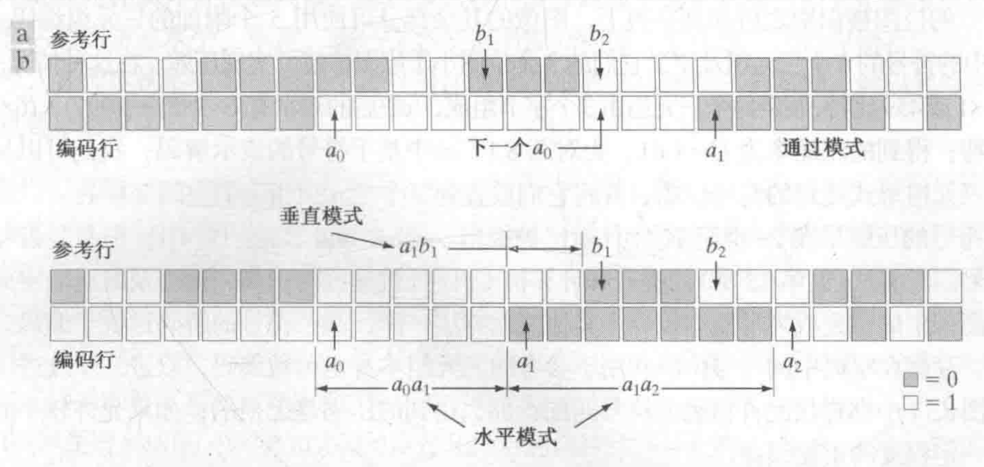


图 8.15　CCITT (a)通过模式和(b)垂直与水平编码模式编码参数

a b

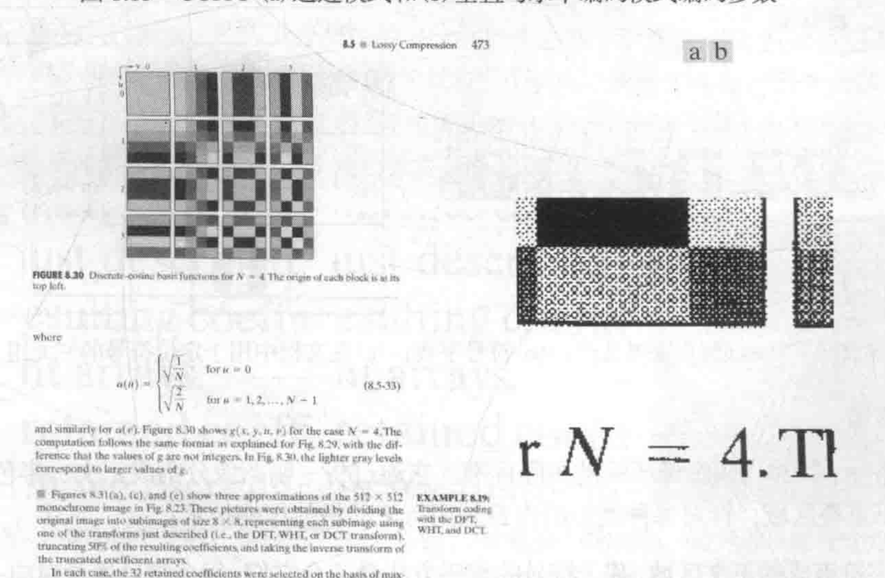

8.5 ■ Lossy Compression　473

FIGURE 8.30 Discrete-cosine basis functions for $N = 4$. The origin of each block is at its top left.

where

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } u = 0 \\ \sqrt{\frac{2}{N}} & \text{for } u = 1, 2, \ldots, N-1 \end{cases} \quad (8.5\text{-}33)$$

and similarly for $\alpha(v)$. Figure 8.30 shows $g(x, y, u, v)$ for the case $N = 4$. The computation follows the same format as explained for Fig. 8.29, with the difference that the values of g are not integers. In Fig. 8.30, the lighter gray levels correspond to larger values of g.

EXAMPLE 8.19: Transform coding with the DFT, WHT, and DCT.

■ Figures 8.31(a), (c), and (e) show three approximations of the 512 × 512 monochrome image in Fig. 8.23. These pictures were obtained by dividing the original image into subimages of size 8 × 8, representing each subimage using one of the transforms just described (i.e., the DFT, WHT, or DCT transform), truncating 50% of the resulting coefficients, and taking the inverse transform of the truncated coefficient arrays.

In each case, the 32 retained coefficients were selected on the basis of maximum magnitude. When we disregard any quantization or coding issues, this process amounts to compressing the original image by a factor of 2. Note that in all cases, the 32 discarded coefficients had little visual impact on reconstructed image quality. Their elimination, however, was accompanied by some mean-square error, which can be seen in the scaled error images of Figs. 8.31(b), (d), and (f). The actual rms errors were 1.28, 0.86, and 0.68 gray levels, respectively. ■

rN = 4.Tl

图 8.16　书页的二值扫描图像：(a)显示整页内容的缩小图像；(b)在抖动中用于显示二值像素的放大图像

8.2.6　基于符号的编码

在基于符号或记号的编码中，图像被表示为多幅频繁发生的子图像的一个集合，称为符号。每个这样的符号都存储在一个符号字典中，且该图像以一个三元组 $\{(x_1, y_1, t_1), (x_2, y_2, t_2), \cdots\}$ 的集合来编码，其中，每个 (x_i, y_j) 对规定了图像中一个符号的位置，而记号 t_i 是该符号或子图像在字典中的地址。也就是说，每个三元组表

> 参考表 8.3 和表 8.4，基于符号的编码用于 JBIG2 压缩中。

581 ~ 582

示图像中一个字典符号的一个实例。通过仅存储一次重复的符号，可以有效地压缩图像，特别是在文档存储和检索应用中，在这种情况下，符号通常是重复多次的字符位图。

考虑图 8.17(a)中简单的两级灰度图像。它包含一个字 *banana*，该字由三个单一符号 a b n 组成：3 个 *a*、1 个 *b* 和 2 个 *n*。假如 *b* 是在编码过程中第一个识别的符号，其 9×7 位图存储在符号字典中的位置 0 内。如图 8.17(b)所示，识别 *b* 位图的记号是 0。这样，在编码后的图像表示中，第一个三元组是(0, 2, 0)［见图 8.17(c)］，它指出表示 *b* 符号的矩形位图的左上角在解码图像中被放在了位置(0, 2)处。对 *a* 和 *n* 符号的位图被识别之后加到字典上，图像的其余部分可使用 5 个附加的三元组编码。只要用于定位图像中的符号的 6 个三元组及定义它们的 3 个位图小于原图像就可实现压缩。在这种情况下，起始图像有 9×51×1 或 459 比特，假定每个三元组由 3 个字节组成，压缩后的表示有(6×3×8) + [(9×7) + (6×7) + (6×6)]或 285 比特；得到的压缩率为 $C = 1.61$。为对图 8.17(c)中基于符号的表示解码，我们可以从符号字典中读取以三元组形式规定的符号位图，并将它们放置到每个三元组指定的空间坐标处。

基于符号的压缩早在 20 世纪 70 年代就已被提出(Ascher and Nagy[1974])，但只是近年来才变得实用。符号匹配算法(见第 12 章)的提升和计算机 CPU 处理速度的提高，使得及时选取字典符号并找到它们在图像中的位置成为可能。像许多其他压缩方法一样，基于符号的解码远快于编码。最后，我们注意到，存储在字典中的符号位图和用于参考的三元组本身也可被编码，以进一步改进压缩性能。如果像在图 8.17 中那样仅允许精确的符号匹配，那么得到的压缩是无损的；如果允许较小的差别，那么将出现一定程度的重建误差。

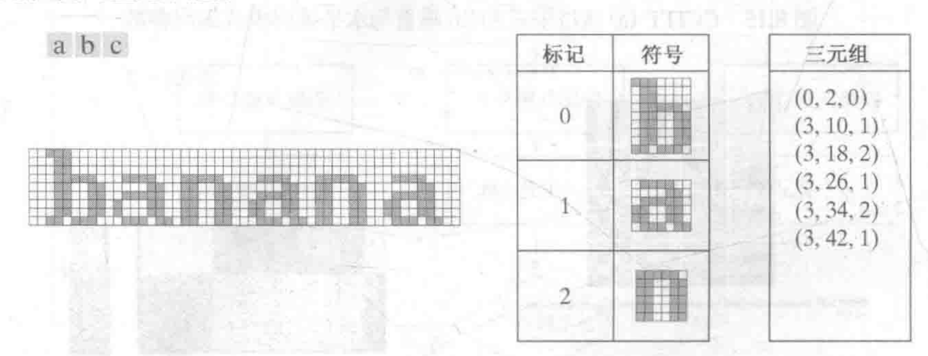


图 8.17 (a)两灰度级文档；(b)符号字典；(c)在文档中用于定位符号的三元组

JBIG2 压缩

JBIG2 是一个针对二级图像压缩的国际标准。它通过将一幅图像分割成正文、半色调和普通内容的重叠和/或不重叠区域，针对每种类型的内容采用特定的最佳压缩技术：

- 对于字符组成的正文区域，基于符号的编码方法最适合它们。每个符号通常对应一个字符位图，即表示一个正文字符的子图像。对于每种采用字体的每个大写和小写字符，在符号字典中通常仅有一个字符位图(或子图像)。例如，在字典中只有一个“a”位图、一个“A”位图、一个“b”位图，等等。

  在有损 JBIG2 压缩中(通常称为*感知无损*或*视觉无损*的 JBIG2)，我们忽略字典位图(即引用字符位图或字符模板)和图像中相应字符的特例之间的差别。在无损压缩中，这一差别被存储，并与三元组联合对每个字符进行编码(由解码器实现)，以产生实际的图像位图。所有的位图不是用算术编码，就是用 MMR(见 8.2.5 节)编码；用于访问字典条目的三元组不是用算术编码，就是用霍夫曼编码。

- 半色调区域类似于正文区域，它们由按规则网格排列的模式组成。然而，存储在字典中的符号

582 ~ 583

不是字符位图，而是表示灰度的周期模式（即一个图形），其已通过抖动来产生用于打印的两灰度级图像。

- 普通区域包含非正文、非半色调信息，类似于工艺线条和噪声，这种区域不是用算术编码方法来压缩，就是用 MMR 编码方法来压缩。

如许多图像压缩标准那样，JBIG2 定义了解码器特性。它确实没有明确地定义一个标准的编码器，但其灵活性足以允许设计各种编码器。虽然编码器的设计处在未确定状态，但它仍然很重要，因为它决定了可能达到的压缩水平。毕竟，编码器必须把图像分割成多个区域，选择存储在字典中的正文和半色调符号，并确定这些符号与图像中符号潜在的情况本质上是相同还是不同。解码器简单地使用这些信息来重建原图像。

例 8.11　JBIG2 压缩实例。

再次考虑图 8.16(a) 中的二级灰度图像。图 8.18(a) 显示了（一个商用文档压缩应用）对该图像进行无损 JBIG2 编码压缩后的一个重建部分。它是原图像的精确复制。注意，在重建文本中，*d* 有轻微变化，尽管它们是由字典中相同的 *d* 条目产生的。*d* 和图像中的 *d* 间的差别被用于改善字典的输出。为在编码字典位图的解码期间实现这一改善，标准定义了一个算法。针对我们的讨论目的，你可以把它想象为将字典位图与图像中对应字符的特定实例间的差别加到从该字典读取的位图上。

图 8.18(b) 是图 8.18(a) 中经感知无损 JBIG2 压缩后的区域的另一个重建。注意，该图中的 *d* 是相同的。它们是直接从符号字典中复制的。称该重建为感知无损的原因是，文本是可读的，甚至字体也相同。我们认为如图 8.18(c) 所示的原图像中的 *d* 和字典中的 *d* 之间的小差别并不重要，因为它们并不影响可读性。记住，我们处理的是二级灰度图像，故图 8.18(c) 中只有三个灰度。灰度 128 指出图 8.18(a) 和图 8.18(b) 中图像对应像素间不存在差别的区域；灰度 0（黑）和灰度 255（白）指出两幅图像中相对灰度的像素，例如，一幅图像中的一个黑像素，在另一幅图像中是白像素，反之亦然。

a b c

just described / esulting coeffic / nt arrays. / retained coeffi / en we disregar

just described / esulting coeffic / nt arrays. / retained coeffi / en we disregar

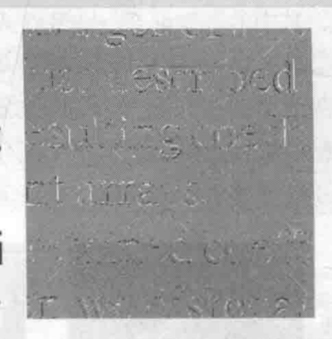


图 8.18　JBIG2 压缩比较：(a) 无损压缩与重建；(b) 感知无损压缩；(c) 两种压缩程度的差值

用于产生图 8.18(a) 的无损 JBIG2 压缩，将无压缩原 PDF 图像的字节数由 663 445 减少到 32 705；压缩率为 $C=20.3$。感知无损 JBIG2 压缩则将原图像的字节数减少到 23 913 字节，压缩率提高到了约 27.7。这些压缩比 CCITT Group 3 和 CCITT Group 4 从例 8.10 得到的结果高 4~5 倍。

8.2.7　比特平面编码

> 参考表 8.3 和表 8.4，比特平面编码用于 JBIG1、JPEG-2000 压缩标准中。

前几节讨论的行程编码技术和基于符号的编码技术，可通过单独处理图像的比特平面的方法用于两级以上灰度的图像。称为比特平面编码的这种技术基于如下概念：把一幅多级（单色或彩色）图像分解为一系列二值图像（见 3.2.4 节），并使用几种熟知的二值压缩方法之一来压缩每幅二值图像。本节描述两种最普遍的分解方法。

583～584

一幅 m 比特单色图像的灰度可以用如下形式的基 2 多项式来表示：

$$a_{m-1}2^{m-1}+a_{m-2}2^{m-2}+\cdots+a_1 2^1+a_0 2^0 \tag{8.2-8}$$

基于这种特性，将该图像分解为二值图像集的一种简单方法是，把该多项式的 m 个系数分离为 m 个 1 比特的比特平面。如 3.2.4 节中解释的那样，最低阶比特平面(对应最低阶比特的平面)是通过收集每个像素的 a_0 比特生成的，而最高阶比特平面包含 a_{m-1} 比特或系数。一般来讲，每个比特平面都由给其像素置一个来自原始图像每一像素的合适比特值或多项式系数来重建。这种分解方法的固有缺点是，灰度的较小变化会对比特平面的复杂性产生明显影响。例如，若一个灰度为 127(01111111)的像素与一个灰度为 128(10000000)的像素相邻，每个比特平面将包含一个对应 0 到 1(或 1 到 0)的转换。譬如，因为 127 和 128 的二进制码的最高阶比特不同，所以最高比特平面将包含一个与值为 1 的像素相邻的零值像素，在该点产生一个 0 到 1(或 1 到 0)的转换。

一种替代分解方法(降低较小灰度变化带来的影响)是，首先用一个 m 比特格雷码表示图像。对应于式(8.2-9)中多项式的 m 比特格雷编码 $g_{m-1}\cdots g_2 g_1 g_0$ 可由下式计算得到：

$$\begin{aligned} g_i &= a_i \oplus a_{i+1},\quad 0\leqslant i\leqslant m-2 \\ g_{m-1} &= a_{m-1} \end{aligned} \tag{8.2-9}$$

式中，⊕ 表示异或运算。这种编码的唯一特性是连续码字只有 1 个比特位不同。因此，较小的灰度变化不太可能影响所有 m 个比特平面。例如，当灰度级 127 和 128 相邻时，只有最高阶比特平面包含有一个 0 到 1 的转换，因为对应于 127 和 128 的格雷码分别是 01000000 和 11000000。

例 8.12　比特平面编码。

图 8.19 和图 8.20 显示了图 8.19(a)中小孩的 8 比特单色图像的 8 个二进制码和格雷码的比特平面。注意，高阶比特平面与其对应的低阶比特平面相比要简单得多。也就是说，这些比特平面中包含有较大的均匀区域，其中的图像细节、杂乱性或随机性都较少。另外，格雷码比特平面比对应的二进制比特平面复杂性要小。表 8.10 中的 JBIG2 编码结果反映了这两种观察。注意，例如，a_5 和 g_5 的压缩结果要比 a_6 和 g_6 压缩的结果大得多；而 g_5 和 g_6 的压缩结果要比 a_5 和 a_6 压缩的结果小。这一倾向贯穿于表的始终，只有 a_0 例外。格雷编码提供了平均约 1.06:1 的压缩优势。结合在一起，格雷编码文件以 678676/475864 或 1.4:1 的比率压缩原始的单色图像；非格雷编码的文件以 678676/503916 或 1.35:1 的比率压缩该图像。



图 8.19　(a)一幅 256 比特单色图像；(b)~(h)图(a)中图像的 4 个最高有效二进制编码和格雷编码比特平面

584 ~ 586

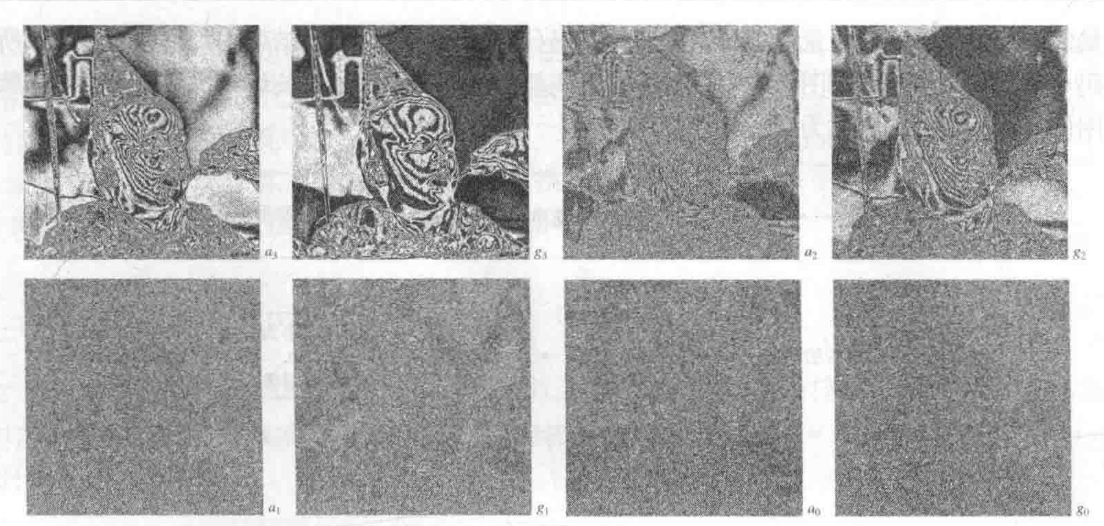


图 8.20　(a)~(h) 图 8.19(a) 中图像的 4 个最低有效二进制编码(左列)和格雷编码(右列)比特平面

表 8.10　图 8.19(a) 的二进制编码和格雷编码比特平面的 JBIG2 无损编码结果

系数 m	二进制编码(PDF，比特)	格雷编码(PDF，比特)	压缩率
7	6 999	6 999	1.00
6	12 791	11 024	1.16
5	40 104	36 914	1.09
4	55 911	47 415	1.18
3	78 915	67 787	1.16
2	101 535	92 630	1.10
1	107 909	105 286	1.03
0	99 753	107 909	0.92

最后，我们注意到图 8.20 中的两个最低阶比特很少有明显的结构。因为这是多数 8 比特单色图像的典型特点，比特平面编码通常限于 6 比特/像素或更少比特/像素的图像。原有的 JBIG1 到 JBIG2 都强加了这样的限制。

8.2.8　块变换编码

本节考虑一种压缩技术，该技术把图像分成大小相等(如 8×8)且不重叠的小块，并使用二维变换单独地处理这些块。在块变换编码中，用一种可逆线性变换(如傅里叶变换)把每个块或子图像映射为变换系数集合，然后对这些变换系数进行量化和编码。对于大多数图像，大量系数都有较小的幅度值，并且可被粗糙地量化(或完全抛弃)而几乎没有多少图像失真。包括第 4 章介绍的离散傅里叶变换在内的各种变换都可用于变换图像数据。

> 参考表 8.3 和表 8.4，块变换编码用于 JPEG、M-JPEG、MPEG-1, 2, 4、H.261, H.262, H.263, H.264、DV 和 HDV、VC-1 及基本压缩标准中。

图8.21 显示了一个典型的块变换编码系统。解码器执行(除了量化功能外)与编码器相反顺序的步骤。编码器执行 4 种相对简单的操作：子图像分解、变换、量化和编码。一幅大小为 $M \times N$ 的输入图像首先被分解为大小为 $n \times n$ 的子图像，然后变换这些子图像以生成 MN/n^2 个子图像变换阵列，每个阵列的大小为 $n \times n$。变换处理的目的，是对每幅子图像中的像素进行去相关运算，或用最少数量的变换系数包含尽可能多的信息。然后，在量化阶段，以一种预定义的方式有选择性地消除或更粗略地量化那些携带最少信息的系数(本节稍后会讨论一些方法)。这些系数对重建的子图像质

> 在本节中，我们重点关注(最常用的)方形子图像。如果需要，我们假设已对输入图像进行了零填充，以便 M 和 N 均是 n 的倍数。

587 ~ 588

量的影响最小。通过对量化后的系数进行编码(通常使用变长编码)结束编码过程。任何或所有变换编码步骤都可以根据局部图像内容进行适应性调整，这称为自适应变换编码，而如果这些步骤对所有子图像都是固定的，则称为非自适应变换编码。

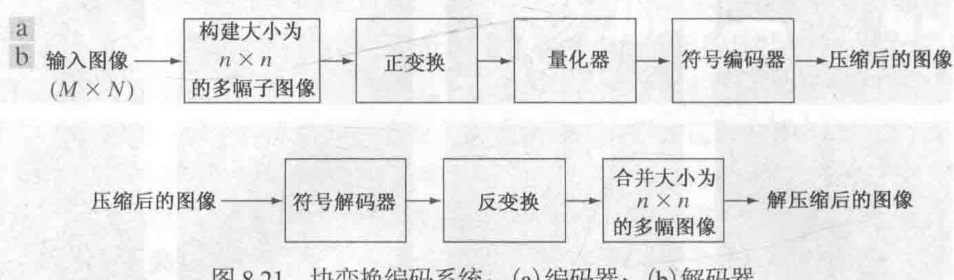


图 8.21　块变换编码系统：(a)编码器；(b)解码器

变换的选择

人们已构建和/或广泛研究了基于各种二维离散变换的块变换编码系统。在给定的应用中，特定变换的选择取决于可容忍的重建误差的大小和可用的计算资源。压缩是在对变换系数的量化期间(而不是变换步骤期间)实现的。

> 我们使用 $g(x, y)$ 来区分来自输入图像 $f(x, y)$ 的子图像。这样，求和限就变成了 n 而不是 M 和 N。

参照 2.6.7 节的讨论，考虑大小为 $n \times n$ 的子图像 $g(x, y)$，其离散变换 $T(u, v)$ 可表示为如下的一般关系：

$$T(u, v) = \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} g(x, y) r(x, y, u, v) \tag{8.2-10}$$

式中，$u, v = 0, 1, 2, \cdots, n-1$。给定 $T(u, v)$，可以使用一般的离散反变换得到 $g(x, y)$：

$$g(x, y) = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} T(u, v) s(x, y, u, v) \tag{8.2-11}$$

式中，$x, y = 0, 1, 2, \cdots, n-1$。在这些公式中，$r(x, y, u, v)$和 $s(x, y, u, v)$分别称为正变换核和反变换核。在本节后面将会很清楚，它们也称为基函数或基图像。式(8.2-10)中的 $T(u, v), u, v = 0, 1, 2, \cdots, n-1$ 称为变换系数；这些系数可视为 $g(x, y)$关于基函数 $s(x, y, u, v)$的一系列展开的展开系数(见 7.2.1 节)。

如 2.6.7 节说明的那样，如果

$$r(x, y, u, v) = r_1(x, u) r_2(y, v) \tag{8.2-12}$$

则式(8.2-10)中的核是可分的；此外，如果 r_1 在函数意义上等于 r_2，则核是对称的。在这种情况下，式(8.2-12)可表示为

$$r(x, y, u, v) = r_1(x, u) r_1(y, v) \tag{8.2-13}$$

如果式(8.2-12)和式(8.2-13)中的 $r(x, y, u, v)$用 $s(x, y, u, v)$代替，则对反变换核也有相同的结论。按 4.11.1 节解释的方式，证明具有可分核的二维变换使用相应一维变换的行-列或列-行操作来计算并不困难。

式(8.2-10)中的正变换核和式(8.2-11)中的反变换核，决定了所计算变换的类型和总体计算的复杂性，以及所采用块变换编码系统的重建误差。最知名的一对变换核是

$$r(x, y, u, v) = \mathrm{e}^{-\mathrm{j}2\pi(ux+vy)/n} \tag{8.2-14}$$

和

$$s(x, y, u, v) = \frac{1}{n^2} \mathrm{e}^{\mathrm{j}2\pi(ux+vy)/n} \tag{8.2-15}$$

588 ~ 590

式中 $j=\sqrt{-1}$。它们是第 2 章中式(2.6-34)和式(2.6-35)在 $M=N=n$ 时定义的变换核。把这些变换核代入式(8.2-10)和式(8.2-11)，可得到 4.5.5 节中介绍的离散傅里叶变换对的一种简化形式。

为计算大小为 $N\times N$ 的输入图像 $f(x,y)$ 而非一幅子图像的 WHT，可在式(8.2-16)中将 n 更改为 N。

变换编码中很有用且计算上较简单的一种变换称为 Walsh-Hadamard 变换(WHT)。这种变换是根据下列功能上相同的核推导而来的：

$$r(x,y,u,v)=s(x,y,u,v)=\frac{1}{n}(-1)^{\sum_{i=0}^{m-1}\left[b_i(x)p_i(u)+b_i(y)p_i(v)\right]} \tag{8.2-16}$$

式中 $n=2^m$。该式中的指数和是以模 2 算术方式执行的，而 $b_k(z)$是 z 的二进制表示中的第 k 位(从右到左)。例如，如果 $m=3$ 且 $z=6$(二进制值为 110)，则 $b_0(z)=0, b_1(z)=1, b_2(z)=1$。式(8.2-16)中的 $p_i(u)$ 使用下式计算：

$$\begin{aligned} p_0(u)&=b_{m-1}(u)\\ p_1(u)&=b_{m-1}(u)+b_{m-2}(u)\\ p_2(u)&=b_{m-2}(u)+b_{m-3}(u)\\ &\vdots\\ p_{m-1}(u)&=b_1(u)+b_0(u) \end{aligned} \tag{8.2-17}$$

如前面提到的那样，式中的求和是以模 2 算术运算执行的。类似的表达式可用于 $p_i(v)$。

不像 DFT 的核，DFT 的核是正弦函数和余弦函数的和[见式(8.2-14)和式(8.2-15)]，Walsh-Hadamard 核由棋盘模式交替排列的+1 和−1 组成。图 8.22 显示了 $n=4$ 时的核。每个块由 $4\times4=16$ 个元素(子方块)组成。白色表示 +1 而黑色表示−1。为得到左上角的块，我们令 $u=v=0$，并画出 $x, y=0, 1, 2, 3$ 时 $r(x,y,0,0)$的值。在这种情况下，所有的值均为+1。顶部行中的第二个块是 $r(x,y,0,1)$在 $x, y=0, 1, 2, 3$ 时的值的图形，以此类推。正如我们已经说明的那样，Walsh-Hadamard 变换的重要性在于它实现上的简单性——所有核的值均为+1 或−1。

图 8.22 $n=4$ 时的 Walsh-Hadamard 基函数。每个块的原点均位于其左上角

为计算大小为 $N\times N$ 的输入图像 $f(x,y)$ 而非一幅子图像的 DCT，可在式(8.2-18)和式(8.2-19)中将 n 更改为 N。

图像压缩中最常使用的变换之一是离散余弦变换(DCT)。这种变换是通过将如下(相等的)核代入式(8.2-10)和式(8.2-11)中得到的：

$$r(x,y,u,v)=s(x,y,u,v)=\alpha(u)\alpha(v)\cos\left[\frac{(2x+1)u\pi}{2n}\right]\cos\left[\frac{(2y+1)v\pi}{2n}\right] \tag{8.2-18}$$

式中，

$$\alpha(u)=\begin{cases}\sqrt{1/n}, & u=0\\ \sqrt{2/n}, & u=1,2,\cdots,n-1\end{cases} \tag{8.2-19}$$

对于 $\alpha(v)$也类似。图 8.23 显示了 $n=4$ 时的 $r(x,y,u,v)$。计算遵循图 8.22 说明的相同格式，不同之处是 r 的值不是整数。在图 8.23 中，较亮的灰度值对应于较大的 r 值。

590 ~ 591

例 8.13 用 DFT、WHT 和 DCT 的块变换编码。

图 8.24(a)到(c)显示了图 8.9(a)中大小为 512×512 的单色图像的三种近似。这些图像是按如下方式得到的：先将原图像分割为大小为 8×8 的子图像，并用刚才描述的变换之一(即 DFT、WHT 或 DCT 变换)来表示每幅子图像，然后截去所得系数的 50%，再对截尾后的系数阵列进行反变换。

在每种情况下，保留下来的 32 个系数都是根据最大幅值选择的。注意，在所有情况下，32 个被丢弃的系数在重建图像质量方面几乎没有视觉影响。然而，它们的消除伴随着一些均方根误差，这一点可在图 8.24(d)到(f)所示的误差图像中看到。实际的均方根误差分别是 2.32, 1.78 和 1.13 个灰度级。

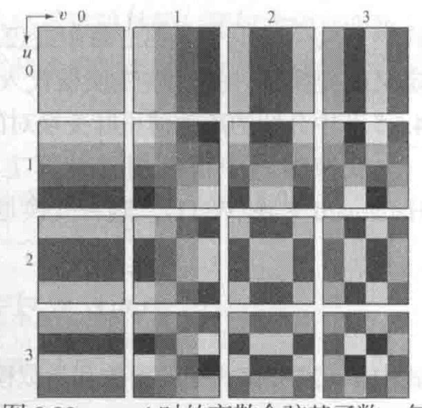


图 8.23 $n=4$ 时的离散余弦基函数。每个块的原点均位于其左上角

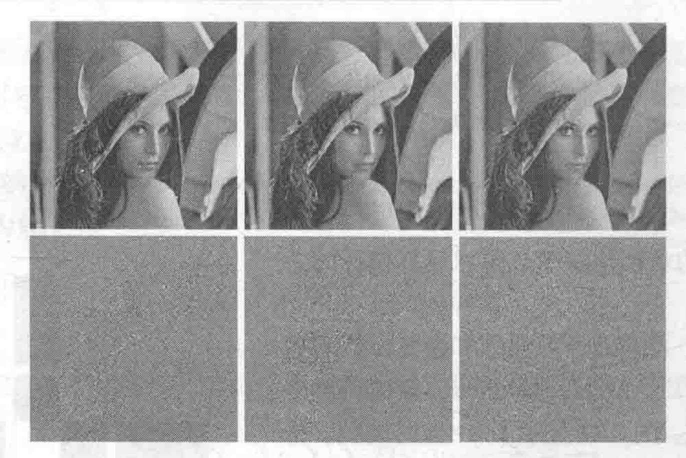


图 8.24 (a)使用傅里叶变换得到的图 8.9(a)的近似；(b)使用 Walshi-Hadamard 变换得到的图8.9(a)的近似；(c)使用余弦变换得到的图8.9(a)的近似；(d)~(f)对应的误差图像

前面的例子中提到的均方重建误差的较小差异，直接与能量或所用变换的信息包装的特性有关。根据式(8.2-11)，大小为 $n\times n$ 的图像 $g(x, y)$可以表示为其二维变换 $T(u, v)$的函数：

$$g(x, y) = \sum_{u=0}^{n-1}\sum_{v=0}^{n-1} T(u, v)s(x, y, u, v) \tag{8.2-20}$$

式中，$x, y = 0, 1, \cdots, 2, n-1$。由于式(8.2-20)中的反变换核 $s(x, y, u, v)$仅取决于参数 x, y, u, v，而不取决于 $g(x, y)$或 $T(u, v)$的值，所以该反变换核可视为由式(8.2-20)定义的序列所定义的一组基函数或基图像。如果式(8.2-20)中所使用的符号进行如下修改，则这种解释会变得很清楚：

$$\boldsymbol{G} = \sum_{u=0}^{n-1}\sum_{v=0}^{n-1} T(u, v)\boldsymbol{S}_{uv} \tag{8.2-21}$$

式中，$\boldsymbol{G}$ 是一个包含 $g(x, y)$的像素的、大小为 $n\times n$ 的矩阵，并且

$$\boldsymbol{S}_{uv} = \begin{bmatrix} s(0, 0, u, v) & s(0, 1, u, v) & \cdots & s(0, n-1, u, v) \\ s(1, 0, u, v) & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & & & \\ s(n-1, 0, u, v) & s(n-1, 1, u, v) & \cdots & s(n-1, n-1, u, v) \end{bmatrix} \tag{8.2-22}$$

591 ~ 593

然后，包含输入子图像像素的矩阵 $\boldsymbol{G}$ 被明确定义为大小为 $n\times n$ 的 n^2 个矩阵的线性组合，即式(8.2-22)中的 $\boldsymbol{S}_{uv}$，$u, v=0,1,2,\cdots,n-1$。实际上，这些矩阵是式(8.2-20)中的级数展开的基图像(或函数)；相关的 $T(u,v)$是展开系数。图 8.22 和图 8.23 图示地说明了 $n=4$ 时的 WHT 和 DCT 基图像。

如果我们现在定义一个变换系数模板函数

$$\chi(u,v)=\begin{cases}0, & \text{若}T(u,v)\text{满足指定的截尾准则}\\ 1, & \text{其他}\end{cases} \tag{8.2-23}$$

式中 $u, v=0,1,2,\cdots,n-1$，则 $\boldsymbol{G}$ 的一个近似可以根据截尾后的展开得到：

$$\hat{\boldsymbol{G}}=\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\chi(u,v)T(u,v)\boldsymbol{S}_{uv} \tag{8.2-24}$$

式中，$\chi(u,v)$ 是用消除公式(8.2-21)中总和贡献最小的基图像构建的。子图像 $\boldsymbol{G}$ 及其近似 $\hat{\boldsymbol{G}}$ 之间的均方误差为

$$\begin{aligned}e_{ms}&=E\left\{\left\|\boldsymbol{G}-\hat{\boldsymbol{G}}\right\|^2\right\}=E\left\{\left\|\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}T(u,v)\boldsymbol{S}_{uv}-\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\chi(u,v)T(u,v)\boldsymbol{S}_{uv}\right\|^2\right\}\\&=E\left\{\left\|\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}T(u,v)\boldsymbol{S}_{uv}\left[1-\chi(u,v)\right]\right\|^2\right\}=\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\sigma_{T(u,v)}^2\left[1-\chi(u,v)\right]\end{aligned} \tag{8.2-25}$$

式中，$\left\|\boldsymbol{G}-\hat{\boldsymbol{G}}\right\|$ 是矩阵 $(\boldsymbol{G}-\hat{\boldsymbol{G}})$ 的范数，$\sigma_{T(u,v)}^2$ 是变换位置(u,v)处系数的方差。最终的化简是以基图像的归一化正交性质和 $\boldsymbol{G}$ 的像素是由零均值随机过程与已知协方差产生的这一假设为基础的。因此，总均方差近似误差是被丢弃的变换系数的方差之和，即 $\chi(u,v)=0$，以便式(8.2-25)中的$[1-\chi(u,v)]$ 为 1 的那些变换系数。将最多的信息重新分配或装入最少的系数中的变换提供最好的子图像近似，从而导致最小的重建误差。最后，在导出式(8.2-25)的假设条件下，一幅大小为 $N\times N$ 的图像的 MN/n^2 个子图像的均方误差是相同的。因此，大小为 $M\times N$ 的图像的均方误差(平均误差的一种度量)等于单个子图像的均方误差。

之前的例子表明，DCT 的信息携带能力要比 DFT 和 WHT 的信息携带能力强。尽管大多数图像通常都拥有这一条件，但在信息携带方面的最佳变换是 Karhunen-Loéve 变换(见第 11 章)而不是 DCT。原因是，对任何输入图像和任何数量的保留系数，KLT 都可以使式(8.2-25)的均方误差最小(Kramer and Mathews[1956])①。然而，由于 KLT 是依赖于数据的，所以为每幅子图像得到 KLT 基图像所需的计算任务一般来说是非常繁重的。由于这个原因，实际上很少使用 KLT 来压缩图像。相反，通常使用像 DFT、WHT 或 DCT 这样具有固定(与输入无关)基图像的变换。与输入无关的变换，非正弦变换(如 WHT 变换)是最容易实现的。正弦变换(如 DFT 或 DCT)的信息携带能力更接近最佳 KLT 的信息携带能力。

因此，多数变换编码系统都基于 DCT，DCT 在信息携带能力和计算复杂性之间提供了较好的折中。事实上，DCT 的性质已经提供了 DCT 成为变换编码系统的国际标准所具有的实际价值。与其他输入独立的变换相比，DCT 变换具有如下优点：用单片集成电路就可以实现，可将最多信息装入最少的系数中②(对大多数图像而言)，并且在子图像间的边界变得可见时，可使出现的称为块缺陷的块效应最小化。与其他正弦变换相比，最后一条性质尤其重要。如图 8.25(a)所示，DFT 隐含的 n 点周期性(见

① 最佳的一个附加条件是式(8.2-23)的模板函数选择具有最大方差的 KLT 系数。

② Ahmed 等人[1974]首先注意到一阶马尔可夫图像源的 KLT 基图像与 DCT 的基图像非常相似。由于相邻像素间的相关接近 1，所以输入相关的 KLT 基图像就变得与输入无关的 DCT 基图像相同(Clarke[1985])。

593 ~ 594

4.6.3 节)造成了导致高频变换内容的边界不连续性的上升。当 DFT 变换系数被截尾或量化时，吉布斯现象[①]会导致边界点呈现错误的值，这些错误的值在图像中以块效应形式出现。也就是说，相邻子图像之间的边界变得可见，因为子图像的边界像素在边界点处呈现为不连续的均值［见图 8.25(a)］。图 8.25(b)的 DCT 减少了这种效应，因为这种变换所隐含的 $2n$ 点周期性本质上不会产生边界不连续。

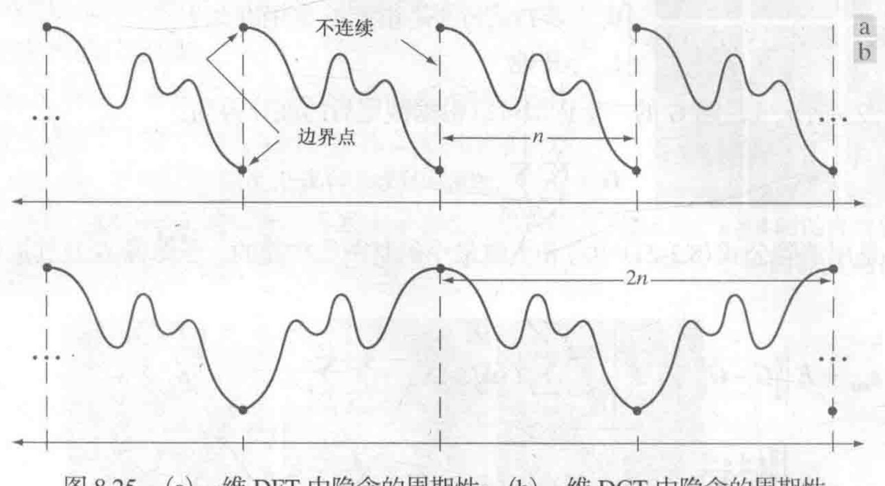


图 8.25 (a)一维 DFT 中隐含的周期性；(b)一维 DCT 中隐含的周期性

子图像尺寸选择

影响变换编码误差和计算复杂性的另一个重要因素是子图像的尺寸。在大多数应用中，图像被进一步细分，以便相邻子图像之间的相关性(冗余)降低到可接受的程度，且 n 为 2 的整数次幂，和前面一样，n 是子图像的维数。后一个条件简化了子图像变换的计算(见 4.11.3 节讨论的基 2 连续加倍方法)。通常，压缩水平和计算复杂性会随子图像尺寸的增加而增大。最常用的子图像尺寸为 8×8 和 16×16。

例 8.14 子图像尺寸对变换编码的影响。

图 8.26 以图形方式说明了子图像尺寸对变换编码重建误差的影响。图中所画的数据是这样得到的：把图8.9(a)中的单色图像分割为大小为 $n\times n$ 的子图像，其中 n = 2, 4, 8, 16, …, 256, 512，计算每幅子图像的变换，截取所得系数的 75%，并对截取后的阵列做反变换。注意，子图像的尺寸大于 8×8 时，Hadamard 和余弦曲线会变得平坦，而傅里叶重建误差在该区域连续减小。n 进一步增大时，傅里叶重建误差将穿过 Walsh-Hadamard 曲线而逼近余弦结果。该结果与由 Netravali and Limb[1980]和 Pratt[1991]报道的二维马尔可夫图像源的理论和实验发现一致。

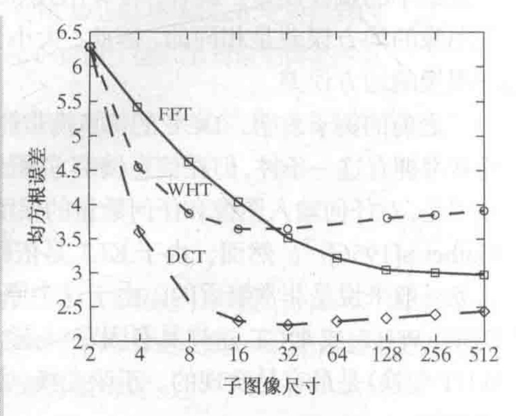
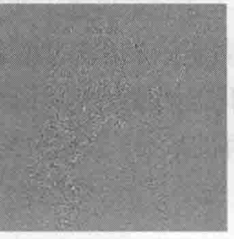


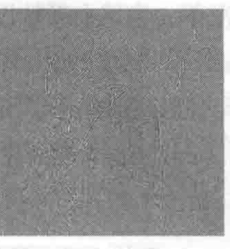
图 8.26 重建误差与子图像尺寸的关系

使用 2×2 大小的子图像时，所有三条曲线相交。在这种情况下，每种变换阵列只有 1/4 的系数(25%)被保留。所有情况下的系数都是直流分量，所以反变换简单地使用它们的均值［见式(4.6-21)］来替代这 4 个子图像像素。这种条件在图 8.27(b)中很明显，这幅图像显示了 2 × 2 大小 DCT 结果放大的一部分。注意块效应，在图 8.27(c)和图 8.27(d)中，随着子图像尺寸增大为 4×4 和 8×8，结果中普遍存在的块效应减少了。作为参考，图 8.27(a)显示了原图像的局部放大。

595 ~ 596

① 多数关于电路分析的电气工程教材中都描述了这种现象，其产生的原因是傅里叶变换在不连续处无法一致收敛。在不连续处，傅里叶展开取不连续点的平均值。



a b c d

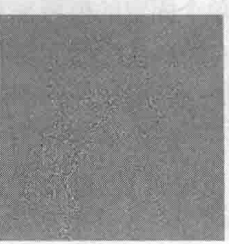
图 8.27　(a)使用 25%的 DCT 系数对图 8.27(a)的近似；(b)使用 2×2 大小的子图的近似；(c)使用 4×4 大小的子图像的近似；(d)使用 8×8 大小的子图像的近似。图(a)中的原图像是图 8.9(a)放大的局部

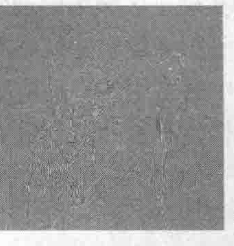
比特分配

与式(8.2-24)截取后的级数展开相关的重建误差，是被丢弃变换系数的数量、相对重要性及用于表示保留系数的精度的函数。在大多数变换编码系统中，保留的系数是根据最大方差进行选择的［即建立式(8.2-23)的模板函数］，这称为区域编码；或是根据最大幅度进行选择的，这称为阈值编码。对变换后的子图像的系数进行截取、量化和编码的整个过程，通常称为比特分配。

例 8.15　比特分配。

图 8.28(a)和(c)显示了图 8.9(a)的两种近似，在这两种近似中，每幅 8×8 子图像丢弃了 87.5%的 DCT 系数。第一幅图像是通过保留 8 个最大的变换系数的阈值编码得到的，第二幅图像是使用区域编码方法生成的。在后一种情况下，每个 DCT 系数都被视为一个随机变量，该随机变量的分布可以通过对所有变换子图像的总体进行计算得到。8 个最大方差的分布(在变换后的 8×8 子图像中，占 64 个系数的 12.5%)被定位并用于确定为所有子图像保留的系数 $T(u, v)$的坐标 u 和 v。注意，图8.28(b)中的阈值编码差值图像所包含的误差比图8.28(d)中区域编码的误差小。为使误差更为明显，两幅图像都已被放大。对应的均方根误差分别是 4.5 和 6.5 个灰度级。



a b c d

图 8.28　用 8×8 DCT 系数的 12.5%对图 8.9(a)的近似：(a)~(b)阈值编码结果；(c)~(d)区域编码结果。放大 4 倍后的差值图像

区域编码的实现　区域编码以信息论中观察信息为不确定性的概念为基础。因此，具有最大方差的变换系数携带大部分图像信息，并在编码处理中应该保留。这些方差本身可以像前面的例子中那样直接根据 MN/n^2 个变换的子图像阵列计算出来，或以一个假设的图像模型为基础(如一个马尔可夫自相关函数)计算出来。在任何一种情况下，根据式(8.2-24)，区域取样处理可被视为每个 $T(u, v)$乘以区域模板中的对应元素，区域模板是在最大方差的位置上设为 1，而在所有其他位置上置 0 构造出来的。最大方差的系数通常被定位在图像变换的原点的周围，所得到的典型区域模板如图 8.29(a)所示。

在区域取样过程中保留的系数必须被量化和编码，以便区域模板有时被描述成用于对每个系数进行编码的比特数［见图 8.29(b)］。在大多数情况下，给这些系数分配相同的比特数，或给这些系数不均等地分配某些固定的比特数。在第一种情况下，系数通常被它们的标准差归一化并被均匀量

597 ~ 598

化。在第二种情况下，为每个系数设计一个量化器，如一个最佳劳埃德-马克斯量化器(见8.2.9节中的最佳量化器)。为构造所要求的量化器，通常用瑞利密度函数对第零个或直流系数进行建模，而剩下的系数则由拉普拉斯或高斯密度函数[①]建模。分配给每个量化器的量化级数(及比特数)与 $\log_2 \sigma^2_{T(u,v)}$ 成正比。这样，式(8.2-24)中保留的系数（基于最大方差选择的系数）就被分配了与系数方差的对数成正比的比特数。

a b
c d

1	1	1	1	1	0	0	0
1	1	1	1	0	0	0	0
1	1	1	0	0	0	0	0
1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

8	7	6	4	3	2	1	0
7	6	5	4	3	2	1	0
6	5	4	3	3	1	1	0
4	4	3	3	2	1	0	0
3	3	3	2	1	1	0	0
2	2	1	1	1	0	0	0
1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	0

1	1	0	1	1	0	0	0
1	1	1	1	0	0	0	0
1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

0	1	5	6	14	15	27	28
2	4	7	13	16	26	29	42
3	8	12	17	25	30	41	43
9	11	18	24	31	40	44	53
10	19	23	32	39	45	52	54
20	22	33	38	46	51	55	60
21	34	37	47	50	56	59	61
35	36	48	49	57	58	62	63

图 8.29 (a)一个典型的区域模板；(b)一个典型的区域比特分配；(c)一个典型的阈值模板；(d)一个典型的经阈值处理后的系数排序序列。用阴影强调的系数是保留的系数

阈值编码的实现 区域编码通常对所有子图像使用单个固定的模板来实现。然而，阈值编码在为每幅彼此不同的子图像保留变换系数的位置方面具有固有的自适应性。实际上，由于其计算的简单性，阈值编码是实际中最常用的自适应变换编码方法。基本概念是对于任何子图像，最大幅值的变换系数对重建后的子图像的质量的贡献最大，正如最后一个例子中展示的那样。

因为不同子图像的最大系数的位置是变化的，所以通常(以事先定义的方式) $\chi(u,v)T(u,v)$ 的元素被重新排序，以形成一个一维行程编码序列。图 8.29(c)显示了一幅假设图像的一幅子图像的一个典型阈值模板。该模板提供了一种对相应子图像进行阈值编码处理及用式(8.2-24)进行数学描述的可视方法。当模板［通过式(8.2-24)］用于产生该模板的子图像，并根据图 8.29(d)中的 Z 形模式对得到的 $n\times n$ 阵列进行重排序，以形成 n^2 个元素的系数序列时，重排序后的一维序列包含有一些 0 的长行程［在图 8.29(d)中，由 0 开始，后跟一系列数字，Z 形模式变得很明显］。这些行程通常被行程编码。对应于包含 1 的模板位置的非零系数或保留系数，通常使用变长码来表示。

有三种基本的途径对一幅变换后的子图像进行阈值处理，或换种说法，有三种基本的途径可以生成式(8.2-23)所示子图像的阈值模板函数：(1)对所有子图像使用单一的全局阈值；(2)对每幅子图像

"最大 N 编码"中的 N，不是一幅图像的维数，而是保留的系数数量。

598
~
599

① 由于每个系数都是其子图像中像素的线性组合［见式(8.2-210)］，中心极限定理告诉我们，当子图像的尺寸增大时，系数就趋向于高斯形式。但是，因为非负图像总有正的直流系数，所以这个结论不适用于直流系数。

使用不同的阈值；(3) 阈值可以随子图像内每个系数的位置的变化而变化。在第一种方法中，对不同图像的压缩水平是不同的，具体取决于超过全局阈值的系数的数量。在第二种方法（被称为最大 N 编码）中，对每幅子图像都丢掉相同数量的系数。结果，代码率是恒定的并且事先已知。在第三种方法中，类似于第一种方法，得到的代码率是变化的，但优点是可以使用下式替代式 (8.2-24) 中的 $\chi(u,v)T(u,v)$，进而把阈值处理和量化结合起来：

$$\hat{T}(u,\ v) = \text{round}\left[\frac{T(u,\ v)}{Z(u,\ v)}\right] \tag{8.2-26}$$

式中，$\hat{T}(u,v)$是 $T(u,v)$经阈值处理和量化后的近似，$Z(u,v)$是变换标准化阵列

$$\boldsymbol{Z} = \begin{bmatrix} Z(0,\ 0) & Z(0,\ 1) & \cdots & Z(0,\ n-1) \\ Z(1,\ 0) & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ Z(n-1,\ 0) & Z(n-1,\ 1) & \cdots & Z(n-1,\ n-1) \end{bmatrix} \tag{8.2-27}$$

的元素。在对子图像变换 $\hat{T}(u,v)$ 标准化(阈值处理和量化)之前，可对其取反变换来得到子图像 $g(x,y)$的一个近似，但 $\hat{T}(u,v)$ 必须乘以 $Z(u,v)$。得到的去标准化阵列表示为 $\dot{T}(u,v)$，它是 $\hat{T}(u,v)$ 的一个近似：

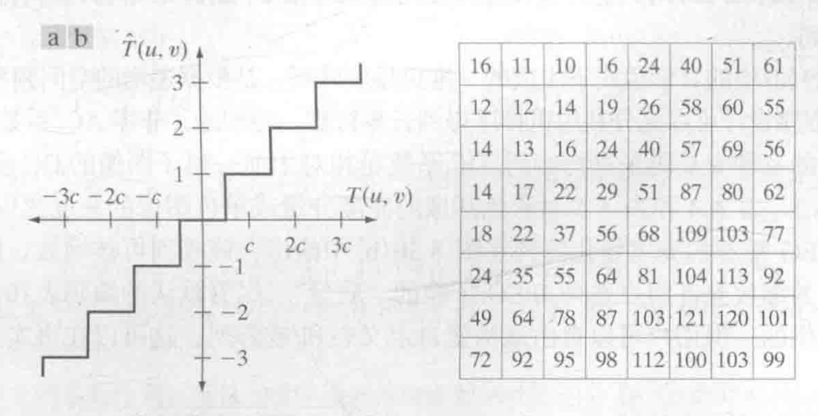
$$\dot{T}(u,\ v) = \hat{T}(u,\ v)Z(u,\ v) \tag{8.2-28}$$

$\dot{T}(u,v)$ 的反变换生成解压缩子图像的近似。

图 8.30 (a) 以图形方式描述了式 (8.2-26) 在 $Z(u,v)$被分配了一个特定值 c 的情况。注意，当且仅当下式成立时，$\hat{T}(u,v)$ 假定整数值 k：

$$kc - \frac{c}{2} \leqslant T(u,\ v) < kc + \frac{c}{2} \tag{8.2-29}$$

如果 $Z(u,v) > 2T(u,v)$，则 $\hat{T}(u,v)=0$，且变换系数被完全截去或丢弃。当 $\hat{T}(u,v)$ 用一种其长度随 k 值的增加而增大的变长编码表示时，用于表示 $T(u,v)$的比特数会受 c 值的控制。因此，$\boldsymbol{Z}$ 的元素可以缩放，以实现各种压缩水平。图8.30 (b) 显示了一个典型的标准化阵列。广泛用于 JPEG 标准(见下节)中的这个阵列，根据试探决定的感知或心理视觉的重要性。对一幅变换后的子图像的每个系数进行加权。



16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

图 8.30　(a) 一条阈值编码量化曲线［见式 (8.2-29)］；(b) 一个典型的标准化矩阵

599
~
601

例 8.16 阈值编码的说明。

图 8.31(a)到(f)显示了 6 种图 8.9(a)中单色图像的阈值编码近似。所有图像都是使用一个 8×8 的 DCT 和图 8.30(b)中的标准化阵列生成的。第一个结果的压缩率约为 12:1(即 $C = 12$)，它是直接应用标准化阵列得到的。其余的结果，分别以 19:1, 30:1, 49:1, 85:1 和 182:1 的压缩率压缩原图，并用 2、4、8、16 和 32 乘以(比例放缩)标准化阵列后生成。相应的均方根误差分别是 3.83, 4.93, 6.62, 9.35, 13.94 和 22.46 个灰度级。



图 8.31 使用 DCT 和图 8.30(b)中的标准化阵列得到的图 8.9(a)的近似：(a) $\boldsymbol{Z}$；(b) $2\boldsymbol{Z}$；(c) $4\boldsymbol{Z}$；(d) $8\boldsymbol{Z}$；(e) $16\boldsymbol{Z}$；(f) $32\boldsymbol{Z}$

JPEG 使用最为普遍且广泛的连续色调静止帧压缩标准是 JPEG 标准。它定义了三种不同的编码系统：(1)一种有损的基本编码系统，该系统以DCT为基础，并且适用于大多数压缩；(2)一种扩展的编码系统，它面向更大压缩、更高精度或渐进的重建应用；(3)一种面向可逆压缩的无损独立编码系统。为实现 JPEG 的兼容性，产品或系统必须包含对基本系统的支持。没有规定特殊的文件格式、空间分辨率或彩色空间模型。

在通常称为顺序基线系统的基线系统中，输入和输出数据的精度限制为 8 比特，而量化后的 DCT 值限制为 11 比特。压缩本身按三个顺序步骤执行：DCT 计算、量化和变长码分配。图像首先被细分为 8×8 的像素块，对这些像素块按从左到右、从上到下的方式进行处理。遇到每个 8×8 的块或子图像时，其 64 个像素通过减去 2^{k-1} 进行灰度级移动，其中 2^k 是灰度级的最大数。然后计算该块的二维离散余弦变换，根据式(8.2-26)对其进行量化，并使用图 8.29(d)的 Z 形模式对其重新排序，形成一个量化系数的一维序列。

由于在图8.29(d)中的 Z 形模式下生成的一维重排序阵列，是根据递增的空间频率定性地排列的，所以 JPEG 编码程序设计可以充分利用重排序得到长零行程。特别地，非零 AC[①]系数使用定义系数值和前面零的个数的一种变长码来进行编码。DC 系数是相对于前一幅子图像的 DC 系数的差值编码。附录 A 中的表 A.3、表 A.4 和表 A.5 对彩色图像的亮度分量或单色图像的灰度提供了默认的 JPEG 霍夫曼编码。JPEG 推荐的亮度量化阵列在图 8.30(b)中给出，该阵列可被缩放，以提供不同的压缩水平。该阵列的缩放允许用户选择 JPEG 压缩的“质量”。尽管默认的编码表和量化阵列是为彩色和单色处理提供的，但用户可以自由地构建自定义表和/或阵列，这可以在事实上适应被压缩图像的特点。

601 ~ 602

① 在该标准中，术语 AC 表示除第零个或 DC 系数外的所有变换系数。

例 8.17　JPEG 基本编码和解码。

考虑下列 8×8 子图像使用 JPEG 基本标准的压缩和重建：

52	55	61	66	70	61	64	73
63	59	66	90	109	85	69	72
62	59	68	113	144	104	66	73
63	58	71	122	154	106	70	69
67	61	68	104	126	88	68	70
79	65	60	70	77	68	58	75
85	71	64	59	55	61	65	83
87	79	69	68	65	76	78	94

原图像由 256 个或 2^8 个可能的灰度组成，因此编码处理由对原始子图像的像素移位 -2^7 或−128 个灰度级开始。得到的移位阵列为

−76	−73	−67	−62	−58	−67	−64	−55
−65	−69	−62	−38	−19	−43	−59	−56
−66	−69	−60	−15	−16	−24	−62	−55
−65	−70	−57	−6	−26	−22	−58	−59
−61	−67	−60	−24	−2	−40	−60	−58
−49	−63	−68	−58	−51	−65	−70	−53
−43	−57	−64	−69	−73	−67	−63	−45
−41	−49	−59	−60	−63	−52	−50	−34

对于 $n = 8$，根据式(8.2-10)和式(8.2-18)的正向 DCT，变换的阵列成为

−415	−29	−62	25	55	−20	−1	3
7	−21	−62	9	11	−7	−6	6
−46	8	77	−25	−30	10	7	−5
−50	13	35	−15	−9	6	0	3
11	−8	−13	−2	−1	1	−4	1
−10	1	3	−3	−1	0	2	−1
−4	−1	2	−1	2	−3	1	−2
−1	−1	−1	−2	−1	−1	0	−1

如果使用图 8.30(b)中 JPEG 推荐的标准阵列来量化这个变换阵列，则缩放和截短［即根据式(8.2-26)进行标准化］后的系数是

−26	−3	−6	2	2	0	0	0
1	−2	−4	0	0	0	0	0
−3	1	5	−1	−1	0	0	0
−4	1	2	−1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

其中，例如，DC 系数计算如下：

$$\hat{T}(0,\ 0) = \text{round}\left[\frac{T(0,\ 0)}{Z(0,\ 0)}\right] = \text{round}\left[\frac{-415}{16}\right] = -26$$

注意，变换和标准化处理产生了大量的零值系数。当根据图 8.29(d)中的 Z 形排序模式对这些系数重新排序时，得到的一维系数序列为

[−26 −3 1 −3 −2 −6 2 −4 1 −4 1 1 5 0 2 0 0 −1 2 0 0 0 0 0 −1 −1 EOB]

其中，符号 EOB 表示块结尾状态。专用的 EOB 霍夫曼码字(见表 A.5 中的分类 0 和行程 0)表明重排序后的序列中的剩余系数均为零。

对于重排序后的系数序列，默认 JPEG 编码的构造是从计算当前 DC 系数和前面已编码子图像的 DC 系数之间的差值开始的。假定紧挨着此子图像左边的变换和量化后的子图像的 DC 系数为−17，则得到的 DPCM

602
~
604

差值为[−26−(−17)]或−9，此值位于表 A.3 的 DC 差值的类别 4 中。根据表 A.4 的默认霍夫曼差值编码，类别 4 差值的正确基础码为 101(一个 3 比特编码)，一个类别 4 的系数的完整编码总长为 7 比特。余下的 4 比特必须从差值的最低阶比特(LSB)生成。对于一个一般的 DC 差值类(譬如类别 K)，需要额外的 K 比特，并且它不是计算正差值的 K 个最低阶比特，就是计算负差值的 K 个最低阶比特减 1。对于差值−9，合适的最低阶比特(LSB)为(0111) − 1 或 0110，而且完全 DPCM 编码的 DC 码字是 1010110。

重排阵列的非零 AC 系数根据表 A.3 和表 A.5 进行类似的编码。主要差别在于，每个默认的 AC 霍夫曼码字依赖于前一个被编码的非零值系数的数量，以及非零系数大小的分类(见表 A.5 中标为“行程/分类”列)。因此，重排阵列的第一个非零 AC 系数(−3)被编码为 0100。该码的前两个比特表明，这个系数属于大小分类 2，并且前面没有零值系数(见表 A.3)；最后两个比特是使用与获得 DC 差值编码的最低阶比特(LSB)相同的处理生成的。继续使用这种方法，完整编码(重排序)阵列为

1010110 0100 001 0100 0101 100001 0110 100011 001

100011 001 001 100101 11100110 110110 0110 11110100 000 1010

为阅读方便起见，这里插入了空格。尽管在这个例子中并不需要这个阵列，但默认的 JPEG 码包含了一个对于 15 个零后跟一个 0 的行程的专用码字(见表 A.5 中的分类 0 和行程 F)。在完全编码后的重排阵列中，总比特数(在这个例子中，要求用于表示整幅 8×8 大小的 8 比特子图像的比特数)是 92。得到的压缩率是 512/92，或大约为 5.6:1。

为解压缩一幅压缩后的 JPEG 子图像，解码器必须首先重建产生压缩比特流的已规格化的变换系数。因为一个经霍夫曼编码的二进制序列是瞬时和唯一可解码的，所以使用简单的查找表方式就可以很容易地完成这一步骤。这里，重新生成的量化系数阵列是：

−26	−2	−6	2	2	0	0	0
1	−2	−4	0	0	0	0	0
−3	1	5	−1	−1	0	0	0
−4	1	2	−1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

根据式(8.2-28)去规格化后，该阵列变成

−416	−33	−60	32	48	0	0	0
12	−24	−56	0	0	0	0	0
−42	13	80	−24	−40	0	0	0
−56	17	44	−29	0	0	0	0
18	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

其中，例如，DC 系数可根据下式计算：

$$\dot{T}(0,\ 0)=\hat{T}(0,\ 0)Z(0,\ 0)=(-26)(16)=-416$$

完全重建的子图像是通过(8.2-11)和式(8.2-18)取去规格化阵列的反 DCT 得到的，由此得到

−70	−64	−61	−64	−69	−66	−58	−50
−72	−73	−61	−39	−30	−40	−54	−59
−68	−78	−58	−9	13	−12	−48	−64
−59	−77	−57	0	22	−13	−51	−60
−54	−75	−64	−23	−13	−44	−63	−60
−52	−71	−72	−54	−54	−71	−71	−54
−45	−59	−70	−68	−67	−67	−61	−50
−35	−47	−61	−66	−60	−48	−44	−44

604 ~ 605

然后对每个反变换后的像素移位 $+2^7$（或+128）个灰度级，得到

58	64	67	64	59	62	70	78
56	55	67	89	98	88	74	69
60	50	70	119	141	116	80	64
69	51	71	128	149	115	77	68
74	53	64	105	115	84	65	72
76	57	56	74	75	57	57	74
83	69	59	60	61	61	67	78
93	81	67	62	69	80	84	84

原始子图像和重建子图像之间的任何差别，都是由 JPEG 压缩和解压缩处理的有损性质造成的。在这个例子中，误差范围是从–14 到+11，并且按如下方式分布：

–6	–9	–6	2	11	–1	–6	–5
7	4	–1	1	11	–3	–5	3
2	9	–2	–6	–3	–12	–14	9
–6	7	0	–4	–5	–9	–7	1
–7	8	4	–1	6	4	3	–2
3	8	4	–4	2	6	1	1
2	2	5	–1	–6	0	–2	5
–6	–2	2	6	–4	–4	–6	10

整个压缩和重建处理的均方根误差接近 5.8 个灰度级。

例 8.18　JPEG 编码的说明。

图 8.32(a)和(d)显示了图 8.9(a)所示单色图像的两种 JPEG 近似。第一个结果提供了 25:1 的压缩；第二个用 52:1 的比率压缩原始图像。图 8.32(a)中的原图像和图 8.32(d)中的重建图像之间的差分别显示在图 8.32(b)和图 8.32(e)中。相应的均方根误差是 5.4 和 10.7 个灰度级。误差在图 8.32(c)和图 8.32(f)的放大图像中清晰可见。这些图像分别显示了图 8.32(a)和图 8.32(d)的局部放大。注意，JPEG 块效应随压缩率的增大而增大。

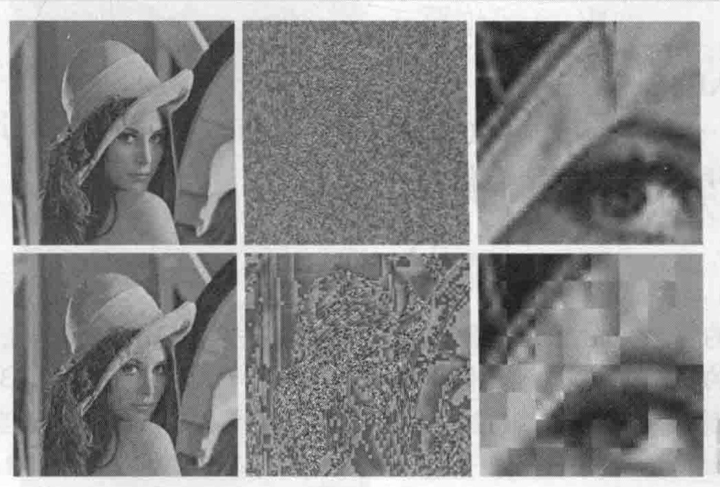


图 8.32　图8.9(a)的两种 JPEG 近似。每一行包含压缩和重建后的结果，该结果和原图像间放大后的差，以及重建图像的局部放大

8.2.9 预测编码

现在我们转向一种更简单的压缩方法，这种方法不需要较大计算开销就可实现较好的压缩效果，并且可以是无误差的或有损的压缩。这种方法通常称为预测编码，它是通过消除紧邻像素在

605 ~ 606

空间和时间上的冗余来实现的，它仅对每个像素中的新信息进行提取和编码。一个像素的新信息定义为该像素的实际值与预测值之间的差。

参考表 8.3 和表 8.4，预测编码用于 JBIG2、JPEG、JPEG-LS、MPEG-1, 2, 4、H.261, H.262, H.263, H.264、HDV、VC-1 和其他压缩标准与文件格式中。

无损预测编码

图8.33显示了一个无损预测编码系统的基本组成。该系统由一个编码器和一个解码器组成，编码器和解码器中均包含有一个相同的预测器。离散时间输入信号 $f(n)$ 的连续样本被传入编码器，预测器根据指定数量的以往样本来生成每个样本的预期值。然后，预测器的输出被四舍五入为最接近的整数，表示为 $\hat{f}(n)$，并使用这个整数来形成差值或预测误差

$$e(n) = f(n) - \hat{f}(n) \tag{8.2-30}$$

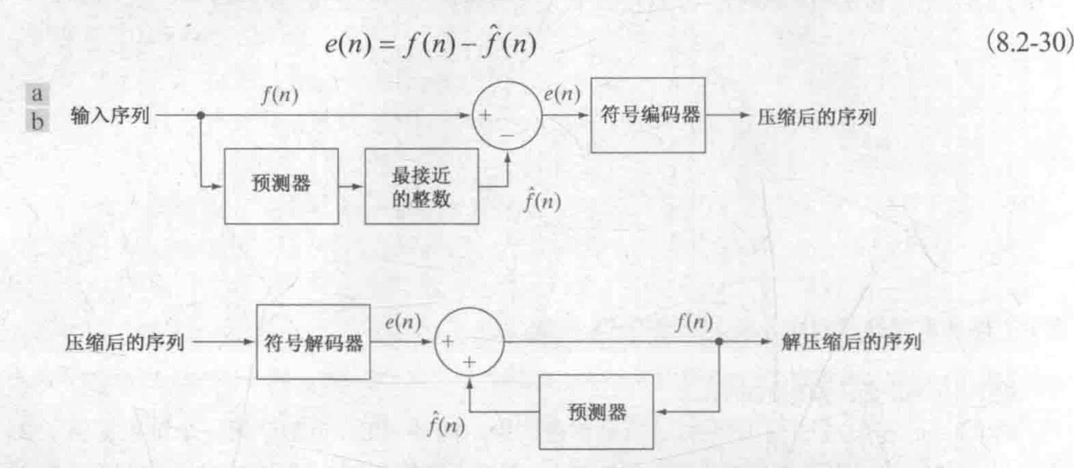


图 8.33 一个无损预测编码模型：(a)编码器；(b)解码器

使用变长码来对这个误差进行编码(用符号编码器)，以生成压缩数据流的下一个元素。图 8.33(b)中的解码器根据接收到的变长码字重建 $e(n)$，并执行反操作

$$f(n) = e(n) + \hat{f}(n) \tag{8.2-31}$$

以解压缩或重建原始输入序列。

可使用各种局部的、全局的和自适应的方法(见"有损预测编码"一节)来生成 $\hat{f}(n)$。在大多数情况下，预测是由前 m 个样值的线性组合生成的，即

$$\hat{f}(n) = \text{round}\left[\sum_{i=1}^{m} \alpha_i f(n-i)\right] \tag{8.2-32}$$

式中，m 是线性预测器的阶，round 表示四舍五入或取最接近整数的操作，α_i，$i = 1, 2, \cdots, m$ 是预测系数。如果将图8.33(a)中的输入序列视为一幅图像的样本，那么式(8.2-30)到式(8.2-32)中的 $f(n)$ 就是像素，用于预测每个像素值的 m 个样本来自当前扫描行(称为一维线性预测编码)、当前行和前几个扫描行(称为二维线性预测编码)，或来自图像序列的当前图像和前几幅图像(称为三维线性预测编码)。这样，对于一维线性预测图像编码，式(8.2-32)可以写为

$$\hat{f}(x, y) = \text{round}\left[\sum_{i=1}^{m} \alpha_i f(x, y-i)\right] \tag{8.2-33}$$

606 ~ 608

其中，每个样本现在被明确地表达为输入图像空间坐标 x 和 y 的函数。注意，式(8.2-33)表明一维线性预测仅是当前行上前几个像素的函数。在二维预测编码中，预测是一幅图像从左到右、从上

到下扫描的前几个像素的函数。在三维情况下，预测是以这些像素和前几帧图像的前几个像素为基础的。式(8.2-33)不能对每一行的前 m 个像素进行求值，因此必须使用其他方式(如霍夫曼编码)来对这些像素编码，并将这种编码考虑为预测编码处理的一种额外开销。类似的说明适用于更高维的情况。

例 8.19　预测编码和空间冗余。

考虑使用来自式(8.2-33)的简单一阶(即 $m=1$)线性预测器对图 8.34(a)中的单色图像进行编码：

$$\hat{f}(x, y) = \text{round}\left[\alpha f(x, y-1)\right] \tag{8.2-34}$$

该式是式(8.2-33)在 $m=1$ 且只有一个预测系数 α_1 (此时下标已无意义)时的简化形式。这种一般形式的预测器称为前像素预测器，对应的预测编码过程称为*差分编码*或*前像素编码*。图 8.34(c)显示了由式(8.2-34)在 $\alpha=1$ 时得到的预测误差图像，即 $e(x,y)=f(x,y)-\hat{f}(x,y)$。这幅图像是按如下方式缩放的：灰度 128 表示一个零预测误差，而所有非零的正预测误差和负预测误差(低于和超过估计)分别显示为更亮的和更暗的灰度色调。预测图像的均值为128.26。因为灰度 128 对应于 0 预测误差，所以平均预测误差仅为 0.26 比特。

图 8.34(b)和(d)分别显示了图 8.34(a)中图像的灰度直方图和预测误差 $e(x,y)$的直方图。注意，图 8.34(d)中预测误差的标准差要比原图像中灰度的标准差小得多。另外，用式(8.1-7)估计的预测误差的熵明显小于原图像的估计熵(3.99 比特/像素对 7.25 比特/像素)。这种熵的降低反映了大量空间冗余的消除，尽管对于 k 比特图像需要 $(k+1)$ 比特数来准确地表示预测误差序列 $e(x,y)$。通常，预测编码方法的最大压缩率可用表示原图像中每个像素的平均比特数除以预测误差的熵的估计来估算。在这个例子中，任何变长编码过程都可用于 $e(x,y)$ 的编码，但得到的压缩率将限制在约 8/3.99 或 2:1 左右。

> 注意，变长编码的预测误差是压缩后的图像。

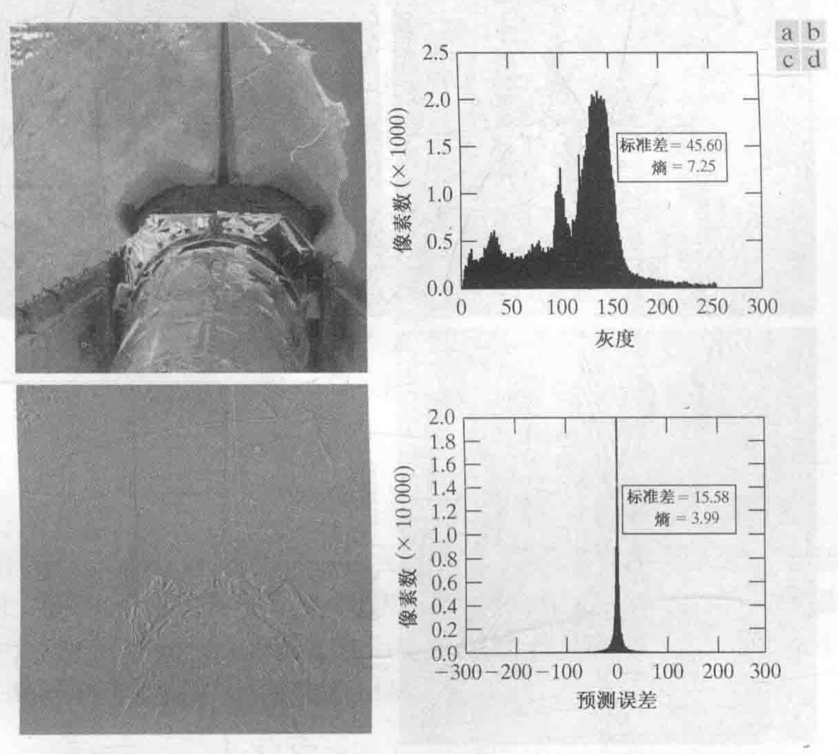


图 8.34　(a)来自轨道空间航天飞机的地球景色；(b)图(a)的灰度直方图；(c)由式(8.2-34)得到的预测误差图像；(d)预测误差的直方图(原图像由 NASA 提供)

608 ~ 609

前面的例子说明，预测编码达到的压缩与将输入图像映射到一个预测误差序列所得到的熵的减少直接相关，通常称为预测残差。因为通过预测和差分处理消除了空间冗余，所以，预测残差的概率密度函数通常在零处会有一个很高的峰值，并由一个相对较小的方差(与输入灰度分布相比)来表征。实际上，它常常由零均值不相关拉普拉斯概率密度函数(PDF)建模：

$$p_e(e) = \frac{1}{\sqrt{2}\sigma_e} e^{\frac{-\sqrt{2}|e|}{\sigma_e}} \tag{8.2-35}$$

式中，σ_e 是 e 的标准差。

例 8.20 预测编码和时间冗余。

图 8.34(a)中的图像是 NASA 的一帧视频的一部分，其中地球正相对于安装在空间航天飞机上的固定摄像机从左到右移动。图 8.35(b)显示的是图 8.35(a)的前一帧的重复图像。使用一阶线性预测器

$$\hat{f}(x, y, t) = \text{round}\left[\alpha f(x, y, t-1)\right] \tag{8.2-36}$$

令 $\alpha = 1$，则图 8.35(b)中的像素的灰度可由图(a)中的对应像素来预测。图 8.34(c)是得到的预测残差图像 $e(x,y,t) = f(x,y,t) - \hat{f}(x,y,t)$。图 8.34(d)是 $e(x,y,t)$ 的直方图。注意，有很小的预测误差。该误差的标准差远小于前一个例子中的标准差——3.76 比特/像素对 15.58 比特/像素。另外，预测误差的熵[使用式(8.1-7)计算]已从 3.99 比特/像素减小到 2.59 比特/像素。通过对得到的预测残差进行变长编码，原图像的压缩率接近 8/2.59 或 3.1∶1——与例 8.19 中使用面向空间的前像素预测器得到的 2∶1 的压缩率相比改进了 50%。

> 再次回忆可知，变长编码后的预测误差是压缩后的图像。

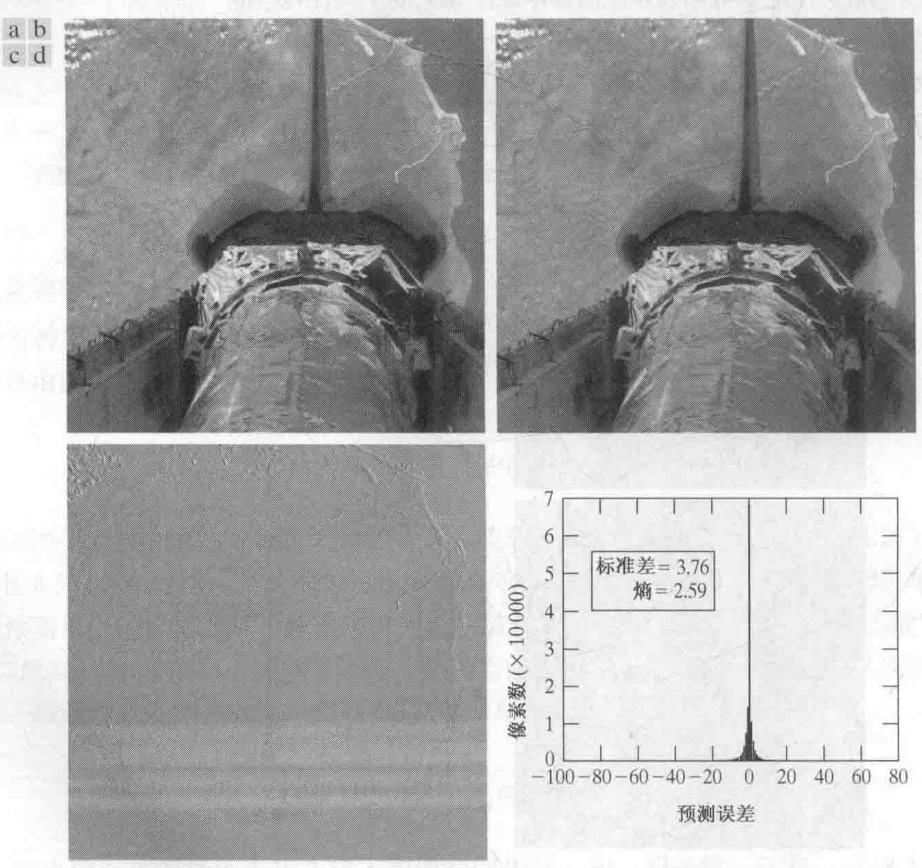


图 8.35 (a)~(b)来自轨道空间航天飞机视频的两幅地球景观图；(c)由式(8.2-36)得到的预测误差图像；(d)预测误差的直方图(原图像由 NASA 提供)

609 ~ 611

运动补偿预测残差

如例 8.20 中看到的那样，视频序列中的连续帧通常是很相似的。对它们的差值进行编码可减少时间冗余并提供有效的压缩。然而，当帧序列包含快速运动的目标时（或者涉及摄像机缩放与摇镜头、场景的突然变化，或者涉及淡入和淡出），相邻帧间的相似性就会降低，从而对压缩率造成负面影响。也就是说，像大多数压缩技术那样(见例8.5)，基于时间的预测编码对某种类型的输入，即具有重要的时间冗余的图像序列，才工作得最好。用于处理有很少时间冗余的图像时，会发生数据扩展。视频压缩系统避免数据扩展问题的方法有如下两种：

1. 在预测和差分处理期间，跟踪目标运动并对其进行补偿。
2. 存在帧间相关(帧间的相似性)而不足以发挥预测编码的优点时，转换到另一种编码方法。

第一种方法(称为运动补偿)是该节剩余内容的主题。但在这之前，我们注意到，当帧间相关不足以使预测编码有效时，第二个问题典型地采用一种面向块的二维变换，如 JPEG 的基于 DCT 的编码(见 8.2.8 节)。在这种方法中，压缩后的帧(即没有预测残差的帧)称为内帧或独立帧(I 帧)。它们可以在不访问属于视频的其他帧的情况下来解码。I 帧通常类似于 JPEG 编码图像，并且是生成预测残差的理想起点。此外，它们提供了高度的随机访问，容易编辑，并能阻止传输误差的传播。因此，所有的标准都要求周期性地将 I 帧插入到压缩的视频码流中。

> "最可能"位置是指这样的一个位置，该位置可使得参考宏块和正被编码宏块间的误差度量最小。这两个块不必是同一目标，但它们必须使误差度量最小。

图 8.36 说明了运动补偿预测编码的基本原理。每个视频帧都被分解成不重叠的矩形区域，这些区域的典型尺寸为 4×4 到 16×16，称为宏块(图 8.36 中仅示出了一个宏块)。每个宏块相对于它在前一视频帧(称为参考帧）中的“最合适的”位置的“运动”，以运动向量来编码。该向量通过定义距“最合适的”位置的水平距离和垂直距离来描述运动。位移通常规定为最近像素、1/2 像素或 1/4 像素精度。如果用亚像素精度，就必须通过参考帧中的像素组合使用内插［如使用双线性内插(见 2.4.4 节)］。一个基于前一帧的编码帧(图 8.36 中的前向预测)称为预测帧(P 帧)；一个基于后一帧的编码帧(图 8.36 的后向预测)称为双向帧(B 帧)。B 帧要求对压缩的码流重新排序，以便这些帧以合适的编码序列而非自然显示顺序的序列提供给解码器。

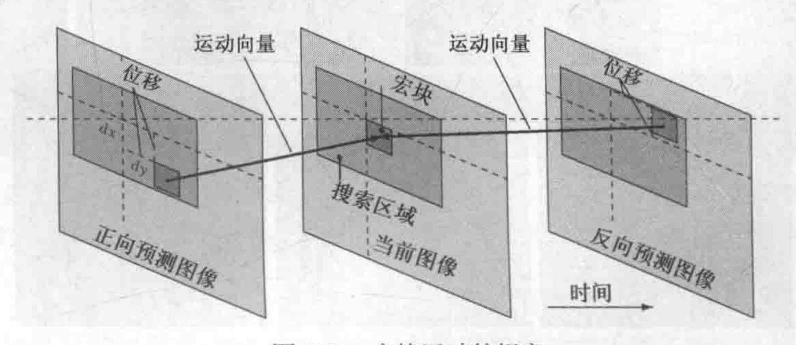


图 8.36　宏块运动的规定

如所期望的那样，运动估计是运动补偿的关键部分。在运动估计期间，物体的运动被测量并编码为运动向量。对“最好的”运动向量的搜索要求定义一个最佳准则。例如，运动向量可在宏块像素和所选参考帧中的预测像素(对于亚向素运动向量内插后的象素)之间的最大相关或最小误差的基础上选择。最常用的一种误差度量是平均绝对失真(MAD)：

$$\mathrm{MAD}(x, y)=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left|f(x+i, y+j)-p(x+i+\mathrm{d}x, y+j+\mathrm{d}y)\right| \tag{8.2-37}$$

611 ~ 612

式中，x 和 y 是被编码的 $m \times n$ 宏块的左上角像素的坐标，dx 和 dy 是如图 8.36 所示的来自参考帧的位移，p 是预测的宏块像素值的一个阵列。对于亚像素运动向量估计，p 来自参考帧中像素的内插。dx 和 dy 通常必须落在围绕每个宏块的一个有限搜索区域内(见图 8.36)，其值通常为 ±8 到 ±64 像素，水平搜索区域通常要比垂直搜索区域稍大。一种更有效的推算误差的测量称为绝对失真和(SAD)，它在式(8.2-37)中省略了因子 $1/mn$。

给定一个如式(8.2-37)所示的选择准则，在运动向量位移允许的范围内，包括亚像素位移，用最小 MAD 搜索 dx 和 dy 执行运动估计。这种处理通常称为块匹配。一种穷举搜索可保证最好可能的结果，但其计算代价高昂，因为必须在整个位移范围内对每个可能的运动进行测试。对于大小为 16×16 的宏块和 ±32 像素的位移范围(对于影片和运动事件没有超出问题的范围)，当使用整数位移精度时，对于一个帧内的每个宏块，必须执行 4225 次 16×16 MAD 计算。如果要求使用 1/2 像素或 1/4 像素的精度，计算的次数还应分别乘以 4 或 16。快速搜索算法可以减少计算负担，但可能会也可能不会得到最佳运动向量。在文献中已经提出并研究了许多基于块的快速运动估计算法(例如，参见 Furht et al.[1997]和 Mitchel et al.[1997])。

例 8.21　运动补偿预测。

图 8.37(a)和(b)取自例 8.19 和例 8.20 中使用过的相同 NASA 视频序列。图 8.37(b)与图 8.34(a)和图8.35(b)相同；图 8.37(a)是之前出现的 13 帧中的一帧的一部分。图 8.37(c)是两帧间的差，它被放大到全部灰度范围。注意，在航天飞机的恒定部分，该差为 0(相对于摄像机)，但在图像的其余部分由于地球的相对运动有较大的差别。图 8.37(c)中预测残差的标准差是 12.73 个灰度级；它的熵［用式(8.1-7)计算］是 4.17 比特/像素。使用变长编码对预测残差进行编码时，可能达到的最大压缩率为 $C = 8/4.17 = 1.92$。

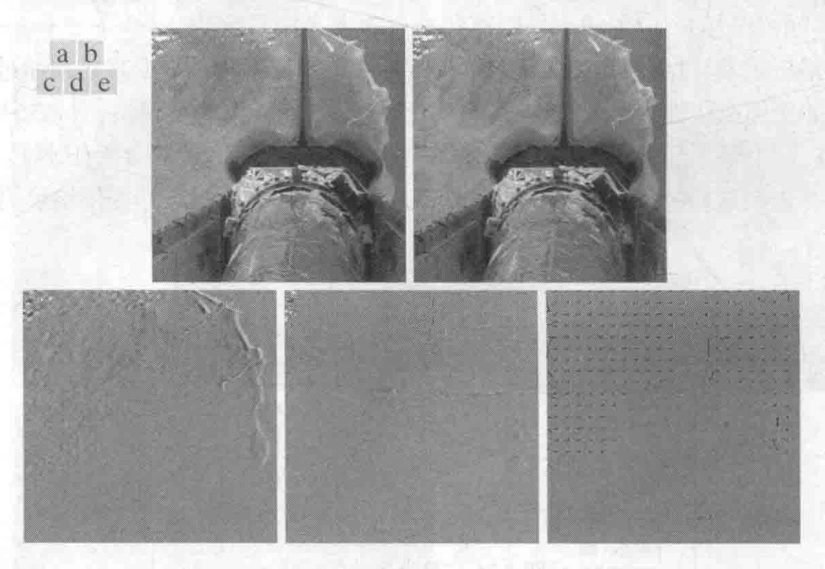


图 8.37　(a)~(b)自绕轨道飞行的航天飞机视频中截取的间隔 13 帧的两幅地球景色图；(c)没有运动补偿的预测误差图像；(d)有运动补偿的预测残差图像；(e)与图(d)相关的运动向量。图(d)中的白点表示描述运动向量的窄头(原图像由 NASA 提供)

图 8.37(d)显示了使用非常低的标准差(5.62 灰度级相对于 12.73 个灰度级)和稍低的熵(3.04 比特/像素相对于 4.17 比特/像素)的运动补偿预测残差。熵用式(8.1-7)计算。如果图 8.37(d)中的预测残差用变长编码，得到的压缩率为 $C = 8/3.04 = 2.63$。为产生这一预测残差，我们把图 8.37(b)分成不重叠的 16×16 宏块，并将每个宏块与图 8.37(a)中的每个 16×16 区域［落入图(b)中宏块位置的 ±16 像素范

612~613

围内的参考帧］相比较。我们使用式(8.2-37)，通过选择具有最低 MAD 的位移(dx, dy)确定最好的匹配。得到的位移是图 8.37(e) 中运动向量的 x 分量和 y 分量。图中的白点是运动向量的头部；它们指出编码宏块的左上角。如从向量模式看到的那样，图像中主要的运动是从左到右的运动。在图像较低的部分（对应于原始图像航天飞机的区域)，由于没有运动，因此没有显示出运动向量。这一区域中的宏块是由参考帧中相似位置的(即相对应的)宏块来预测的。因为图 8.37(e) 中的运动向量是高度相关的，所以可以对它们进行变长编码以减少它们的存储和传输需求。

图8.38说明了可使用亚像素运动补偿提高预测精度。图8.38(a)是图 8.37(c) 的重复，并包括一个参考点；它显示了没有运动补偿的预测误差。图 8.38(b)，(c) 和 (d) 是经运动补偿后的预测残差。它们基于例 8.21 中用过的相同两帧，并使用宏块位移分别为 1, 1/2 和 1/4 像素分辨率(即精度)来计算。使用大小为 8×8 的宏块；位移被限制到 ±8 个像素。

> 图 8.37(c) 和图 8.38(a) 间的视觉差别是由缩放造成的。图 8.38(a) 中的图像已被缩放到与图 8.38(b)~(d) 相匹配。

图 8.38 中预测残差间最重要的视觉差别是灰度的峰值和谷值（即灰度最暗和最亮的区域）的数量与大小。图 8.38(d) 中的 1/4 像素残差是 4 幅图像中“最平坦的”，具有最少的黑或白偏移。如所期望的那样，它有最窄的直方图。图 8.38(a) 到 (d) 中的预测残差的标准差随运动向量精度的增加而减小——分别从 12.7 像素减小到 4.4 像素，从 12.7 像素减小到 4 像素和减小到 3.8 像素。由式(8.1-7)确定的残差的熵分别为 4.17, 3.34, 3.35 和 3.34 比特/像素。这样，运动补偿后的残差就包含相同的信息量，尽管图 8.38(c) 和 (d) 中的残差是使用附加比特来提供 1/2 和 1/4 像素的内插。最后，我们注意到，在每个运动补偿后的残差的左侧，都有一条明显的增大预测误差的条带。这是由于地球从左到右的运动，它把地球上的新的或以前看不到的地形区域带进了每幅图像中的左侧区域。因为前几帧中不存在这些区域，所以不管使用的计算运动向量的精度如何，它们都不能被准确地预测。

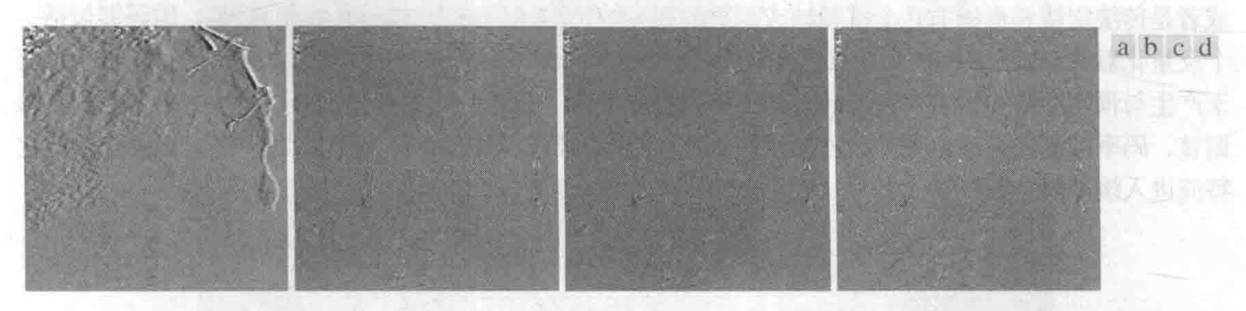


图 8.38　亚像素运动补偿后的预测残差：(a) 没有运动补偿；(b) 单像素精度；(c) 1/2 像素精度；(d) 1/4 像素精度(所有预测误差都已被放大到整个灰度范围，然后乘以 2，以增加它们的可视性)

运动估计是一项要求计算量的任务。所幸的是，只有编码器必须估计宏块运动。给定这些宏块的运动向量，解码器只须简单地访问编码器中用于形成预测残差的参考帧的区域。因为这一原因，运动估计未包括在大多数视频压缩标准中。压缩标准集中在解码器上——在宏块维数、运动向量精度、水平和垂直位移范围等参数上设置约束条件。表 8.11 给出了一些最重要的视频压缩标准的关键预测编码参数。注意，针对 I 帧编码，大多数标准使用大小为 8×8 的 DCT，但对运动补偿规定了更大的区域(如大小为 16×16 的宏块)。另外，由于 DCT 系数量化的有效性，P 帧和 B 帧预测残差甚至也用变换编码。最后，我们注意到，H.261 和 MPEG-4 AVC 标准支持帧内预测编码(在 I 帧内)，以减少空间冗余。

613 ~ 616

表 8.11 视频压缩标准中的预测编码

参 数	H.261	MPEG-1	H.262 MPEG-2	H.263	MPEG-4	VC-1 WMV-9	H.264 MPEG-4 AVC
运动向量精度	1	1/2	1/2	1/2	1/4	1/4	1/4
宏块大小	16×16	16×16	16×16 16×8	16×16 8×8	16×16 8×8	16×16 8×8	16×16 16×8 8×16 8×8 8×4 4×8 4×4
变换	8×8 DCT	8×8 DCT	8×8 DCT	8×8 DCT	8×8 DCT	8×8 8×4 4×8 4×4 整数 DCT	4×4 8×8 整数
帧间预测	P	P, B	P, B	P, B	P, B	P, B	P, B
I 帧帧内预测	否	否	否	否	否	否	是

> 按本章之前的定义，量化器是不可逆的。图 8.39 中的“反量化器”确实不会阻止信息损失。

图 8.39 显示了一个典型的运动补偿视频编码器。它利用了帧内和相邻视频帧间的冗余度、帧间的运动均匀性及人类视觉系统的心理视觉特性。我们可以认为输入到编码器的是顺序的视频宏块。对于彩色视频，每个宏块都由一个亮度块和两个彩色块组成。因为眼睛对彩色的空间敏锐度远小于对亮度的空间敏锐度，所以彩色块通常以亮度块的水平和垂直分辨率的一半来取样。图中加灰的部分是 JPEG 编码器的并行变换、量化和变长编码操作。主要的差别是输入，它可能是图像数据的传统宏块(对于 I 帧)，或者是传统宏块和根据前几个视频帧或相继的视频帧预测间的差值(对于 P 帧和 B 帧)。编码器包括一个反量化器和反映射器(即反 DCT)，以便其预测与互补解码器的那些预测相匹配。此外，它被设计用于产生与预期的视频信道容量相匹配的压缩比特流。为完成这一任务，量化参数由一个码率控制器来调整，码率控制器是输出缓冲器占用量的函数。当缓冲器变得较满时，量化就粗一些，以便较少的比特流进入缓冲器。

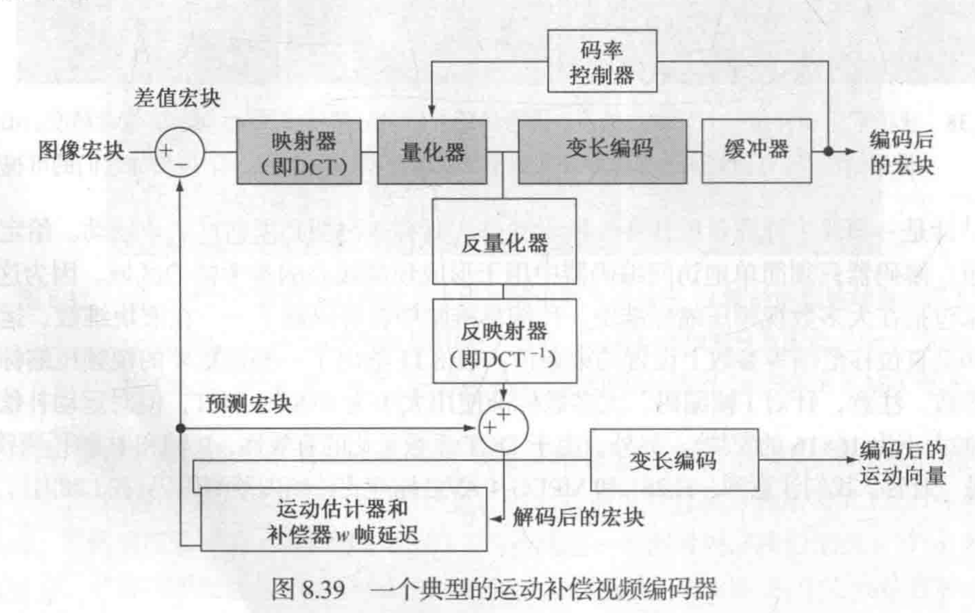


图 8.39 一个典型的运动补偿视频编码器

616 ~ 617

例 8.22 视频压缩的例子。

我们用一个例子来结束对运动补偿预测编码的讨论，该例说明了使用现代视频压缩方法的压缩种类。图 8.40 显示了 1 分钟的 NASA 高清(1280×720)全彩色视频的 15 帧，它们的一部分已在本节用过。虽然显示的图像是单色的，但视频是 1829 幅全彩色帧序列。注意，有各种场景、大量的运动和多样的淡入淡出效果。例如，该视频的 150 帧以黑色淡入方式打开，包括图 8.40 中的第 21 帧和第 44 帧，并用图 8.40 中包含的第 1595 帧、第 1609 帧和第 1652 帧的淡出序列，并紧接着淡出到黑色结束。还有一些急剧变化的场景，如图 8.40 中第 1303 帧和第 1304 帧所涉及的变化。

本节中使用的 NASA 视频片断，请参阅本书的网站。

以 Quicktime 文件(见表 8.4)存储的 NASA 视频 H.264 压缩的版本要求 44.56 MB 的存储空间——加上另外 1.39MB 的相关音频。视频质量很高。对于不压缩的全彩色图像视频帧的数据存储需要约为 5 GB。还应注意到，视频还包含涉及旋转和缩放变化的序列(即图 8.40 中包括第 959 帧、第 1023 帧和第 1088 帧的序列)，但本节的讨论仅限于平移。

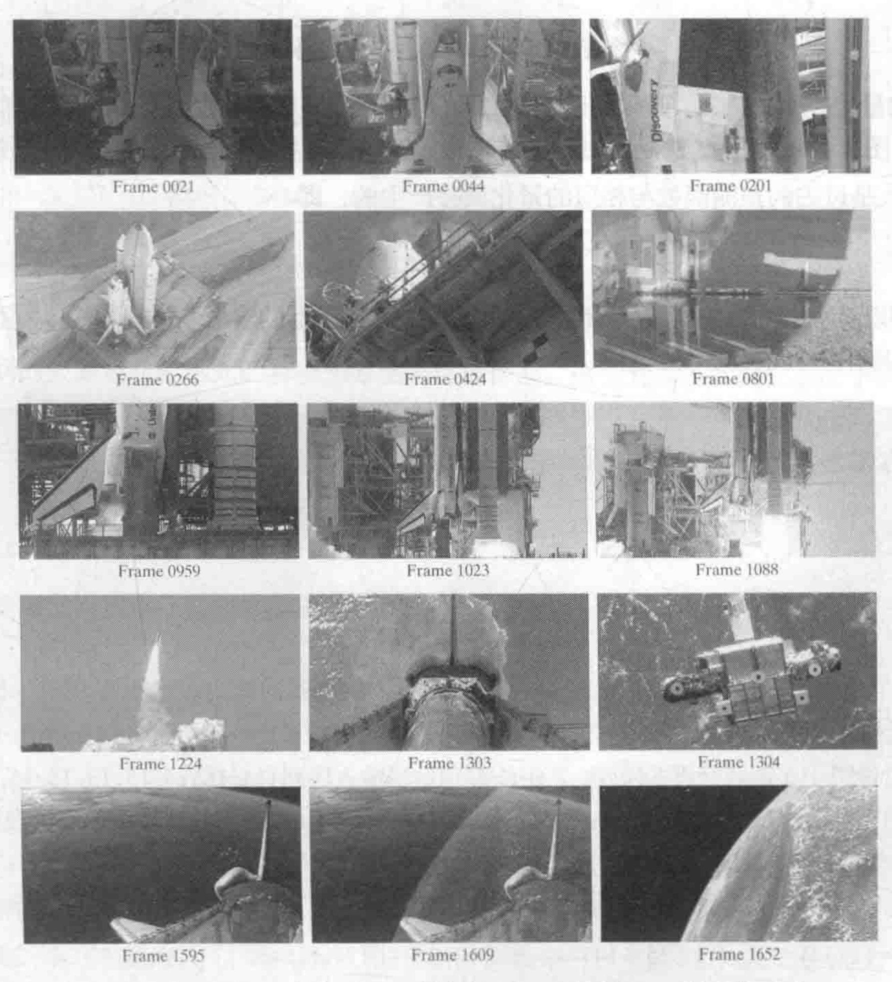


图 8.40 来自 1 分钟、1829 帧的 NASA 视频的 15 帧图像。原始视频是全彩色高清视频(原视频由 NASA 提供)

有损预测编码

本节在先前介绍的无损预测编码模型中加入一个量化器，并在关于空间预测器的上下文中，探讨重建精度和压缩性能间的折中。如图 8.41 所显示的那样，代替无误差编码器的取最接近整数

617 ~ 618

功能的量化器，被插入到了符号编码器和形成预测误差的那一点之间。该量化器将预测误差映射为有限范围内的输出，表示为$\dot{e}(n)$，它确定了压缩量和产生的失真量。

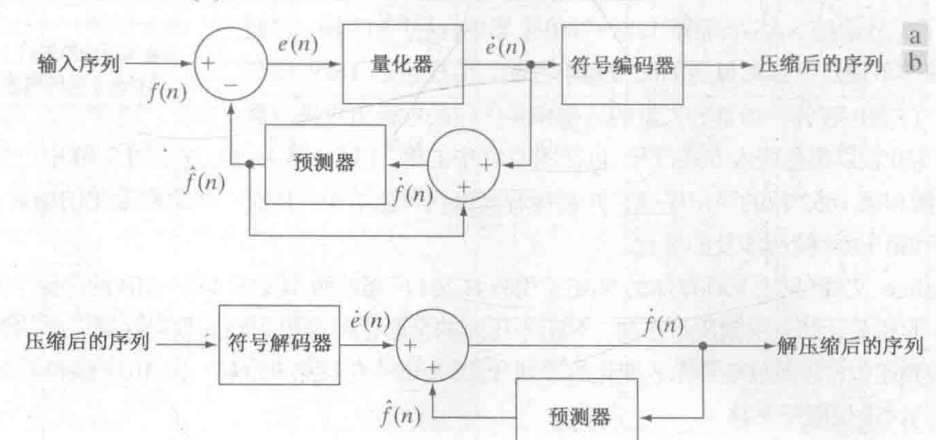


图 8.41　一个无损预测编码模型：(a) 编码器；(b) 解码器

为了接纳量化步骤的加入，必须更改图 8.33(a) 中的无误差编码器，以便由编码器和解码器产生的预测相等。如图 8.41(a) 所示，这是通过在反馈环中放置一个有损编码器的预测器来完成的，其中表示为$\dot{f}(n)$的输入是过去的预测函数与相应的量化误差产生的，即

$$\dot{f}(n)=\dot{e}(n)+\hat{f}(n) \tag{8.2-38}$$

式中，$\hat{f}(n)$同以前所定义的一样。这个闭环结构可以防止在解码器的输出处形成误差。注意，在图 8.41(b) 中，解码器的输出也由式(8.2-38)给出。

例 8.23　Δ 调制。

Δ 调制(DM)是一种简单而知名的有损预测编码形式，其中预测器和量化器定义为

$$\hat{f}(n)=\alpha\dot{f}(n-1) \tag{8.2-39}$$

和

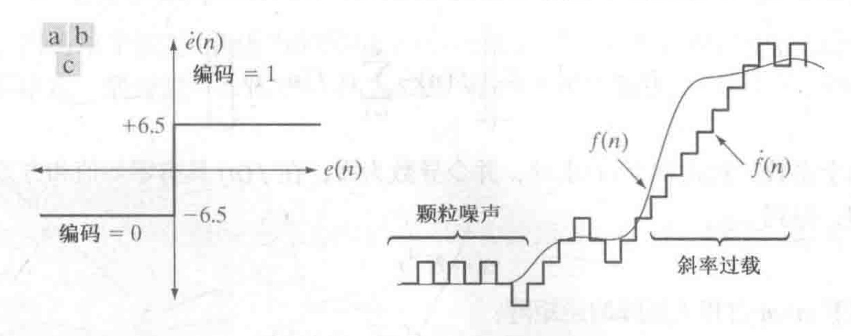
$$\dot{e}(n)=\begin{cases}+\zeta, & e(n)>0\\ -\zeta, & 其他\end{cases} \tag{8.2-40}$$

式中，α是一个预测系数(通常小于 1)，ζ是一个正常量。量化器的输出$\dot{e}(n)$可以用单一比特表示［见图 8.42(a)］，因此，图 8.41(a) 中的符号编码器可使用 1 比特的定长码。得到的 DM 码率为 1 比特/像素。

图 8.42(c) 说明了 Δ 调制处理的机理，其中压缩和重建输入序列{14, 15, 14, 15, 13, 15, 15, 14, 20, 26, 27, 28, 27, 27, 29, 37, 47, 62, 75, 77, 78, 79, 80, 81, 81, 82, 82}, $\alpha=1$, $\zeta=6.5$ 所需的计算被编成了表格。该处理过程从第一个样本无误差地输入到解码器中开始。编码器和解码器处设立的初始条件为$\dot{f}(0)=f(0)=14$，可以通过反复用式(8.2-39)、式(8.2-30)、式(8.2-40)和式(8.2-38)计算出其余的输出。这样，例如，当$n=1$时，就有$\hat{f}(1)=(1)(14)=14$, $e(1)=15-14=1$, $\dot{e}(1)=+6.5$［因为$e(1)>0$］, $\dot{f}(1)=6.4+14=20.5$，得到的重建误差是$(15-20.5)$或-5.5。

图8.42(b) 以图形方式显示了图8.42(c) 中的图表化数据。显示了输入和完全解码后的输出［即$f(n)$和$\dot{f}(n)$］。注意，在从$n=14$到 19 的这一快速变换区域中，ζ太小而不足以表示输入的最大变化，产生了称为斜率过载的失真。此外，当ζ太大而无法表示输入的较小变化时，如在从$n=0$到$n=7$的相对平滑的区域中，出现了颗粒噪声。在图像中，这两种现象会导致模糊的物体边缘和颗粒表面或噪声表面(即失真的平滑区域)。

619 ~ 620



输入		编码器				解码器		误差
n	$f(n)$	$\hat{f}(n)$	$e(n)$	$\dot{e}(n)$	$\dot{f}(n)$	$\hat{f}(n)$	$\dot{f}(n)$	$f(n)-\dot{f}(n)$
0	14	—	—	—	14.0	—	14.0	0.0
1	15	14.0	1.0	6.5	20.5	14.0	20.5	−5.5
2	14	20.5	−6.5	−6.5	14.0	20.5	14.0	0.0
3	15	14.0	1.0	6.5	20.5	14.0	20.5	−5.5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
14	29	20.5	8.5	6.5	27.0	20.5	27.0	2.0
15	37	27.0	10.0	6.5	33.5	27.0	33.5	3.5
16	47	33.5	13.5	6.5	40.0	33.5	40.0	7.0
17	62	40.0	22.0	6.5	46.5	40.0	46.5	15.5
18	75	46.5	28.5	6.5	53.0	46.5	53.0	22.0
19	77	53.0	24.0	6.5	59.6	53.0	59.6	17.5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

图 8.42　Δ 调制的一个例子

前例中提到的失真对所有形式的有损预测编码都很普遍。这些失真的严重性取决于所用量化和预测方法之间的一个相互作用的复杂集合。先不管它们相互间的作用，预测器通常是在无量化误差的假设之下设计的，并且量化器设计为使其自身的误差最小。也就是说，预测器和量化器是彼此独立进行设计的。

最佳预测器

在多数预测编码应用中，选择使编码器的均方预测误差最小的预测器

$$E\left\{e^2(n)\right\} = E\left\{\left[f(n)-\hat{f}(n)\right]^2\right\} \tag{8.2-41}$$

约束条件[①]为

$$\dot{f}(n) = \dot{e}(n)+\hat{f}(n) \approx e(n)+\hat{f}(n) = f(n) \tag{8.2-42}$$

和

$$\hat{f}(n) = \sum_{i=1}^{m} \alpha_i f(n-i) \tag{8.2-43}$$

也就是说，最佳的准则是最小均方预测误差，量化误差假设是可以忽略的 [$\dot{e}(n) \approx e(n)$]，且预测被约束为前 m 个样本的一个线性组合[②]。这些约束并不是本质的，但它们在相当程度上简化了分析，同时降低了预测器计算的复杂性。如此得到的预测编码方法称为*差分脉冲编码调制*(DPCM)。

在这些条件下，最佳预测器设计问题就简化成了相对简单地选择 m 个预测系数，这些系数使表达式

620 ~ 621

① 符号 $E\{\}$ 代表统计期望算子。

② 一般来说，一个非高斯序列的最佳预测器，是用于形成估计的样本的一个非线性函数。

$$E\left\{e^2(n)\right\} = E\left\{\left[f(n) - \sum_{i=1}^{m} \alpha_i f(n-i)\right]^2\right\} \tag{8.2-44}$$

最小。相对于每个系数，对式(8.2-44)求导，并令导数为零，在 $f(n)$ 具有零均值和方差 σ^2 的假设下，求联立方程的解，得到

$$\boldsymbol{\alpha} = \boldsymbol{R}^{-1}\boldsymbol{r} \tag{8.2-45}$$

式中，$\boldsymbol{R}^{-1}$ 是如下 $m\times m$ 自相关矩阵的逆矩阵：

$$\boldsymbol{R} = \begin{bmatrix} E\{f(n-1)f(n-1)\} & E\{f(n-1)f(n-2)\} & \cdots & E\{f(n-1)f(n-m)\} \\ E\{f(n-2)f(n-1)\} & \vdots & \cdots & \vdots \\ \vdots & & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & & & \\ E\{f(n-m)f(n-1)\} & E\{f(n-m)f(n-2)\} & \cdots & E\{f(n-m)f(n-m)\} \end{bmatrix} \tag{8.2-46}$$

$\boldsymbol{r}$ 和 $\boldsymbol{\alpha}$ 是 m 元向量：

$$\boldsymbol{r} = \begin{bmatrix} E\{f(n)f(n-1)\} \\ E\{f(n)f(n-2)\} \\ \vdots \\ E\{f(n)f(n-m)\} \end{bmatrix} \text{和 } \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix} \tag{8.2-47}$$

因此，对于任何输入序列，使式(8.2-44)最小的系数可以通过一系列基本矩阵操作来确定。此外，这些系数仅取决于原序列中样本的自相关。使用这些最佳系数得到的预测误差的方差为

$$\sigma_e^2 = \sigma_2 - \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{r} = \sigma^2 - \sum_{i=1}^{m} E\{f(n)f(n-i)\}\alpha_i \tag{8.2-48}$$

尽管计算式(8.2-45)的机理非常简单，但形成 $\boldsymbol{R}$ 和 $\boldsymbol{r}$ 所需的自相关计算实现起来却非常困难，因此，在实践中几乎从不使用局部预测(即在这些预测中对每个输入序列计算预测系数)。在大多数情况下，通过假设一个简单的输入模型，并将相应的自相关代入式(8.2-46)和式(8.2-47)(见8.1.4节)来计算一组全局系数。譬如，假设带有可离分自相关函数的二维马尔可夫图像源(见8.1.4节)

$$E\{f(x,y)f(x-i,y-j)\} = \sigma^2 \rho_v^i \rho_h^j \tag{8.2-49}$$

和广义四阶线性预测器

$$\hat{f}(x,y) = \alpha_1 f(x,y-1) + \alpha_2 f(x-1,y-1) + \alpha_3 f(x-1,y) + \alpha_4 f(x-1,y+1) \tag{8.2-50}$$

得到的最佳系数是

$$\alpha_1 = \rho_h,\ \alpha_2 = -\rho_v\rho_h,\ \alpha_3 = \rho_v,\ \alpha_4 = 0 \tag{8.2-51}$$

式中，ρ_h 和 ρ_v 分别是所研究图像的水平和垂直相关系数。

最后，通常要求式(8.2-43)中的预测系数之和小于等于1，即

$$\sum_{i=1}^{m} \alpha_i \leqslant 1 \tag{8.2-52}$$

621 ~ 622

这种限制是为了确保预测器的输出落入允许输入范围内，并减少传输噪声的影响[当输入到图8.41(a)中的是一幅图像时，传输噪声在重建图像中通常表现为水平条纹]。降低 DPCM 解码器对输入噪声的

敏感性很重要，因为单个误差(在适当的环境下)会传播到以后的所有输出中，也就是说，解码器的输出会变得不稳定。通过进一步限制式(8.2-52)严格小于 1，可将输入误差的影响限制在少量的输出中。

例 8.24　预测技术的比较。

考虑由对图 8.9(a)中的单色图像进行 DPCM 编码产生的预测误差，假设量化误差为零且使用如下 4 个预测器：

$$\hat{f}(x, y) = 0.97 f(x, y-1) \tag{8.2-53}$$

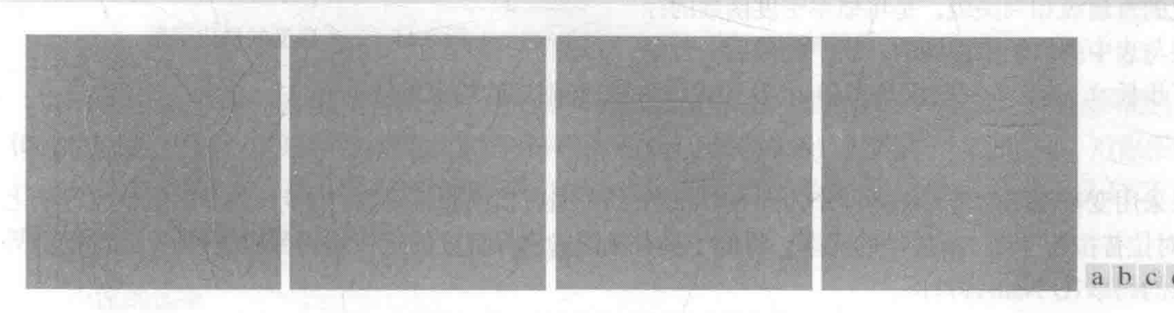
$$\hat{f}(x, y) = 0.5 f(x, y-1) + 0.5 f(x-1, y) \tag{8.2-54}$$

$$\hat{f}(x, y) = 0.75 f(x, y-1) + 0.75 f(x-1, y) - 0.5 f(x-1, y-1) \tag{8.2-55}$$

$$\hat{f}(x,y) = \begin{cases} 0.97 f(x,y-1), & \Delta h \leqslant \Delta v \\ 0.97 f(x-1,y), & \text{其他} \end{cases} \tag{8.2-56}$$

式中，$\Delta h = |f(x-1,y) - f(x-1,y-1)|$ 和 $\Delta v = |f(x,y-1) - f(x-1,y-1)|$ 表示在点(x, y)处的水平和垂直梯度。式(8.2-53)到式(8.2-56)定义了 α_i 的一个相对鲁棒的集合，该集合可对很宽范围的图像提供令人满意的性能。设计式(8.2-56)的自适应预测器的目的是，计算一幅图像的方向特性的局部度量(Δh 和 Δv)，并选择适合特定度量性能的预测器来改善边缘的重现质量。

图8.43(a)到(d)显示了使用式(8.2-53)到式(8.2-56)的预测器得到的预测误差图像。注意，随着预测器阶数的增加明显，可觉察的误差减少了[①]。预测误差的标准差遵循类似的模式，它们分别是 11.1, 9.8, 9.1 和 9.7 个灰度级。



a b c d

图 8.43　4 种线性预测技术的比较

最佳量化

图8.44中的阶梯量化函数 $t = q(s)$ 是 s 的奇函数［即 $q(-s) = -q(s)$］，这个函数完全可以由图8.44第一象限中所示的 s_i 和 t_i 的 $L/2$ 个值来描述。这些断点定义了函数的不连续性，并称为该量化器的判决级和重建级。为方便起见，如果 s 位于半开区间 $(s_i, s_{i+1}]$ 内，则可以将 s 视为到 t_i 的映射。

量化器设计问题是对于一种特殊的最佳准则和输入概率密度函数 $p(s)$，选择最好的 s_i 和 t_i。如果最佳准则(可以是统计度量或心理视觉度量[②])是均方量化误差［即 $E\{(s_i - t_i)^2\}$］最小，并且 $p(s)$ 是一个偶函数，则最小误差(Max[1960])的条件是

$$\int_{s_i-1}^{s_i} (s - t_i) p(s) \mathrm{d}s = 0, \quad i = 1, 2, \cdots, \frac{L}{2} \tag{8.2-57}$$

① 使用多于 3 个或 4 个像素的预测器，会提供很少的压缩增益，但复杂性却会增加(Habibi[1971])。

② 关于心理视觉度量的详细内容，请参阅 Netravali[1977]和 Limb and Rubinstein[1978]。

$$s_i=\begin{cases}0, & i=0\\ \dfrac{t_i+t_{i+1}}{2}, & i=1,\ 2,\cdots,L/2-1\\ \infty, & i=L/2\end{cases} \tag{8.2-58}$$

和

$$s_{-i}=-s_i,\ t_{-i}=-t_i \tag{8.2-59}$$

式(8.2-57)指出重建级是规定的判决区间上 $p(s)$ 下方面积的质心，而式(8.2-59)指出判决级是重建级之间的一半。式(8.2-59)是 q 为一个奇函数的结果。对于任意 L，满足式(8.2-57)到式(8.2-59)的 s_i 和 t_i 在均方误差意义下是最佳的；相应的量化器称为 L 级劳埃德-马克斯(Lloyd-Max)量化器。

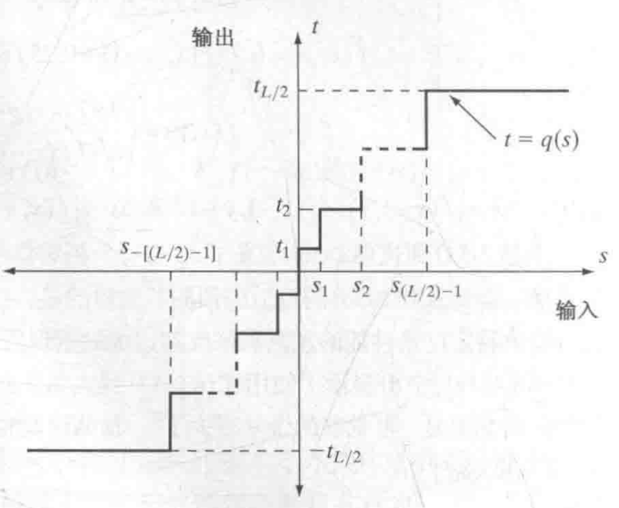


图 8.44 一个典型的量化函数

表8.12列出了单位方差拉普拉斯概率密度函数[见式(8.2-35)]的 2 级、4 级和 8 级劳埃德-马克斯判决和重建级。因为对大多数非平凡的 $p(s)$ 来说，得到式(8.2-57)到式(8.2-59)的显式或闭合形式的解很困难，所以这些值都是数值生成的(Paez and Glisson[1972])。所示的这三个量化器分别提供了 1, 2 和 3 比特/像素的固定输出速率。由于表 8.12 是对单位方差分布构建的，所以 $\sigma\neq 1$ 情况的重建级和判决级，是将概率密度函数的标准差与表中的值相乘得到的。表中的最后一行列出了步长 θ，步长 θ 同时满足式(8.2-57)到式(8.2-59)和附加的约束条件：

$$t_i-t_{i-1}=s_i-s_{i-1}=\theta \tag{8.2-60}$$

如果采用变长编码的符号编码器作为图 8.41(a)中的普通有损预测编码器，则步长为 θ 的最佳均匀量化器(对拉普拉斯 PDF)所提供的码率，将低于具有相同输出保真度的定长编码劳埃德-马克斯量化器所提供的码率(O'Neil[1971])。

表 8.12 单位方差拉普拉斯概率密度函数的劳埃德-马克斯量化器

级别	2		4		8	
i	s_i	t_i	s_i	t_i	s_i	t_i
1	∞	0.707	1.102	0.395	0.504	0.222
2			∞	1.810	1.181	0.785
3					2.285	1.576
4					∞	2.994
θ	1.414		1.087		0.731	

尽管劳埃德-马克斯和最佳均匀量化器不是自适应的，但以图像的局部特性为基础来调整量化级别好处更多。理论上，缓慢变化的区域可被精细量化，而快速变化的区域可被更为粗糙地量化。这种方法可以同时降低颗粒噪声和斜率过载，而在码率上仅要求极小的增加。代价是增加了量化器的复杂性。

624~625

8.2.10 小波编码

类似于8.2.8节的变换编码技术，小波编码基于以下概念：对图像像素解除相关的变换系数进行编码，比对原图像像素本身进行编码的效率更高。如果变换的基函数（此时为小波函数）将大多数重要的可视信息包装到少量系数中，则剩下的系数可被粗略地量化或截取为零，而图像几乎没有失真。

> 参考表 8.3 和表 8.4，小波编码用于 JPEG-2000 压缩标准中。

图8.45显示了一个典型的小波编码系统。为了对一幅大小为 $2^J \times 2^J$ 的图像进行编码，选择一种分析小波 ψ 和最小分解级别 $J-P$，并用于计算图像的离散小波变换。如果小波具有互补的尺度函数 φ，则可以使用快速小波变换(见 7.4 节和 7.5 节)。不论哪种情况，计算出来的变换都会将原图像的大部分转换为水平、垂直和对角分解系数，这些系数具有零均值和类似拉普拉斯分布。回顾图 7.1 中的图像和图 7.10(a)中图像的小波变换的较简单的统计，由于许多计算的系数携带很少的视觉信息，这些系数可以最小的系数和编码冗余来量化和编码。此外，量化可以自适应地越过 P 分解级别而利用任何位置相关。一种或多种无损编码方法，如行程编码、霍夫曼编码、算术编码和比特平面编码等，都可以应用到最后的符号编码步骤中。解码可以用与编码相反的操作来实现——量化过程除外，因为量化过程不可能严格逆向执行。

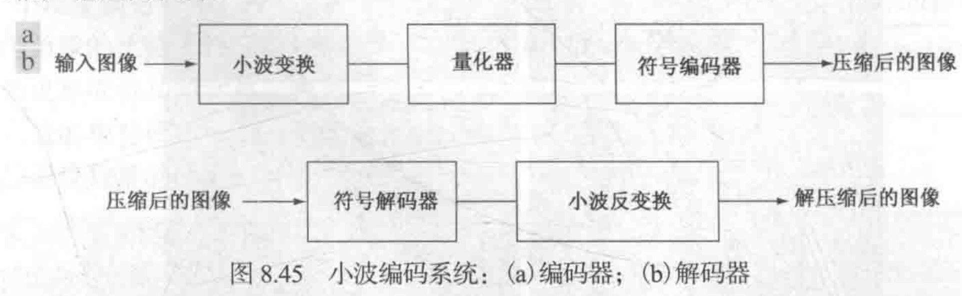


图 8.45 小波编码系统：(a)编码器；(b)解码器

图 8.45 中基于小波的系统与图 8.21 中的变换编码系统之间的主要差别是，省略了变换编码器的子图像处理阶段。因为小波变换的计算效率和固有的局部性(即小波的基函数在宽度上是有限的)，对原图像进行细分是没有必要的。稍后我们会看到，细分步骤的取消可以消除块效应，这种效应正是以 DCT 为基础的近似图像在高压缩比下的特性。

小波的选择

为图 8.45 中正变换和逆变换的基所选择的小波，影响着小波编码系统的设计和性能的各个方面。它们直接影响到变换的计算复杂性，或间接影响压缩和重建具有可接受误差的图像系统的能力。当变换小波带有一个尺度函数时，变换可以通过一系列数字滤波操作来实现。滤波器抽头的数量等于非零小波和尺度向量系数的数量。小波将信息包装到较少变换系数中的能力，取决于小波的压缩和重建性能。

> 在数字滤波中，每个滤波器抽头都会将正被滤波信号的一个延迟形式乘以一个滤波器系数。

基于小波的压缩广泛使用的展开函数是 Daubechies 小波和双正交小波。后者允许将有用的分析特性，如零矩的数量(见7.5节)合并到分解滤波器中，而将重要的综合特性，如重建的平滑度，嵌入到重建滤波器中。

例 8.25 小波编码中的小波基。

图 8.46 包含了图 8.9(a)中的 4 个离散小波变换。Haar(哈尔)小波是此例中考虑的最简单且唯一不连续的小波，这种小波被用做图 8.46(a)中的展开函数或基函数。图 8.46(b)中应用最普遍的成像小波是 Daubechies 小

> DWT 细节系数已在7.3.2 节中讨论过。

626 ~ 627

波和图8.46(c)中使用的对称小波，它们是具有增加对称性的 Daubechies 小波的一种扩展。在图8.46(d)中使用的 Cohen-Daubechies-Feauveau 小波用于说明双正交小波的性能。如以前的这类结果那样，所有细节系数都进行了标定，以便使基本结构更为突出——灰度 128 对应于系数值 0。

如表 8.13 所示，从图 8.46(a)移到图 8.46(d)时，图 8.46 中变换的计算所涉及的操作次数，从每个系数(对于每个分解级别)的 4 次乘法和加法增加到 28 次乘法和加法。所有 4 个变换都使用一种快速小波变换(即滤波器组)公式进行计算。注意，当计算的复杂性(即滤波器抽头的数量)增加时，信息包装性能也会更好。采用哈尔小波且小于 1.5 的细节系数被截短为零时，总变换的 33.8%也都为零。对于更复杂的双正交小波，被置零的系数的数量增加到 42.1%，潜在压缩率提升了近 10%。

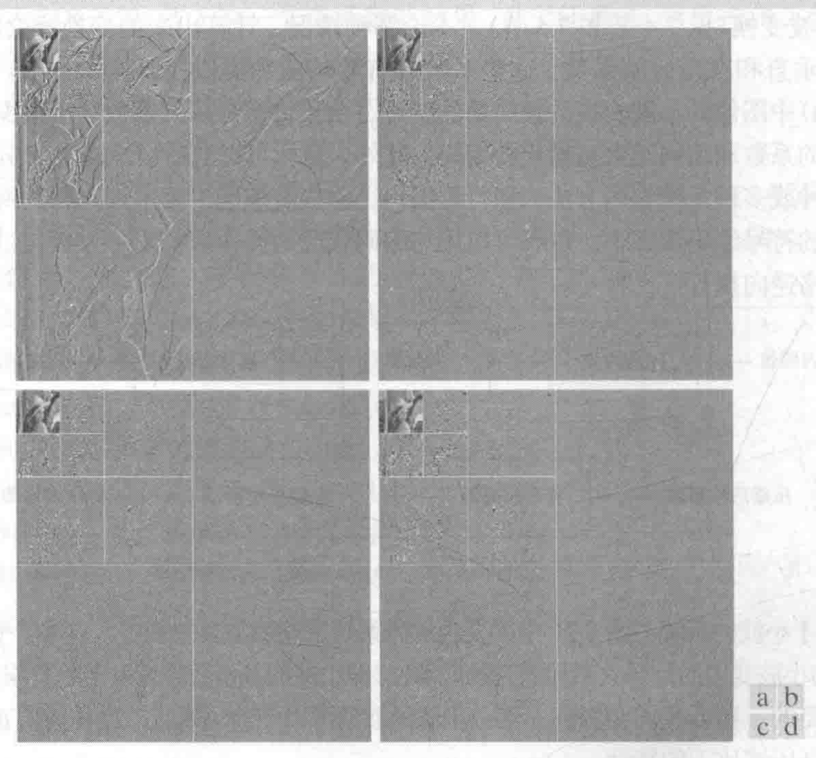


图 8.46 图 8.9(a)关于(a)哈尔小波、(b)Daubechies 小波、(c)对称小波和(d)Cohen-Daubechies-Feauveau 双正交小波的三尺度小波变换

表 8.13 将图 8.46 中细节系数小于 1.5 的变换截短为零时，小波变换滤波器的抽头和置零系数

小 波	滤波器抽头(尺度+小波)	置 零 系 数
哈尔小波(见习题 7.10)	2 + 2	33.8%
Daubechies 小波(见图 7.8)	8 + 8	40.9%
对称小波(见图 7.26)	8 + 8	41.2%
双正交小波(见图 7.39)	17 + 11	42.1%

分解级别选取

另一种影响小波编码计算复杂性和重建误差的因素是变换分解级别的数量。由于 P 尺度快速小波变换涉及 P 个滤波器组的迭代，正变换和反变换计算中的操作次数会随分解级数的增加而增加。此外，对更大分解级别导致的逐渐增大的低尺度系数的量化，会影响重建图像中逐渐增大的区域。在许多应用中，如搜索图像数据库或逐渐重建传送图像等，所存储或传送图像的分辨率和最低可用近似的尺度通常决定了变换级别的数量。

例 8.26　小波编码中的分解级别。

表 8.14 说明了使用双正交小波和一个固定全局阈值 25 对图 8.9(a) 进行编码时，分解级别选择的影响。如之前小波编码的例子中那样，只截短细节系数。表中列出了置零系数的百分比和由式 (8.1-10) 得到的均方根重建误差。注意，初始分解决定多数数据压缩。在 3 个分解级别上，被截短系数的数量几乎没有变化。

表 8.14　对图 8.9(a) 中大小为 512×512 的图像进行小波编码时分解级别的影响

分解级别(尺度或滤波器组迭代)	近似系数图像	截短的系数(%)	重建误差(rms)
1	256×256	74.7%	3.27
2	128×128	91.7%	4.23
3	64×64	95.1%	4.54
4	32×32	95.6%	4.61
5	16×16	95.5%	4.63

量化器设计

影响小波编码压缩和重建误差的最重要因素是系数量化。尽管最广泛使用的量化器是均匀的，但量化的效果可以通过以下方法进一步改进：(1) 引入一个以零为中心的较大量化间隔，称为死区，或 (2) 从一个尺度到另一个尺度自适应调整量化间隔的大小。不论哪种情况，选择的量化间隔都必须随着编码图像的比特流传送给解码器。间隔本身可根据被压缩的图像试探来决定或自动地计算。例如，一个全局系数阈值可作为第一级细节系数的绝对值的中值，或作为被截短的零的数量和重建图像中保留的能量数量的函数来计算。

> 数字信号的能量度量是样本平方之和。

例 8.27　小波编码中死区间隔的选取

图8.47说明了在截短的细节系数的百分比下，死区间隔大小对图 8.9(a) 中基于三尺度双正交小波编码的影响。当死区尺寸增大时，被截短系数的数量同样增大。在曲线拐点的上方（即横坐标 5 处），被截短系数的数量几乎没有再增加。这是因为细节系数的直方图在零的周围有很高的峰值（例如，见图 7.10）。

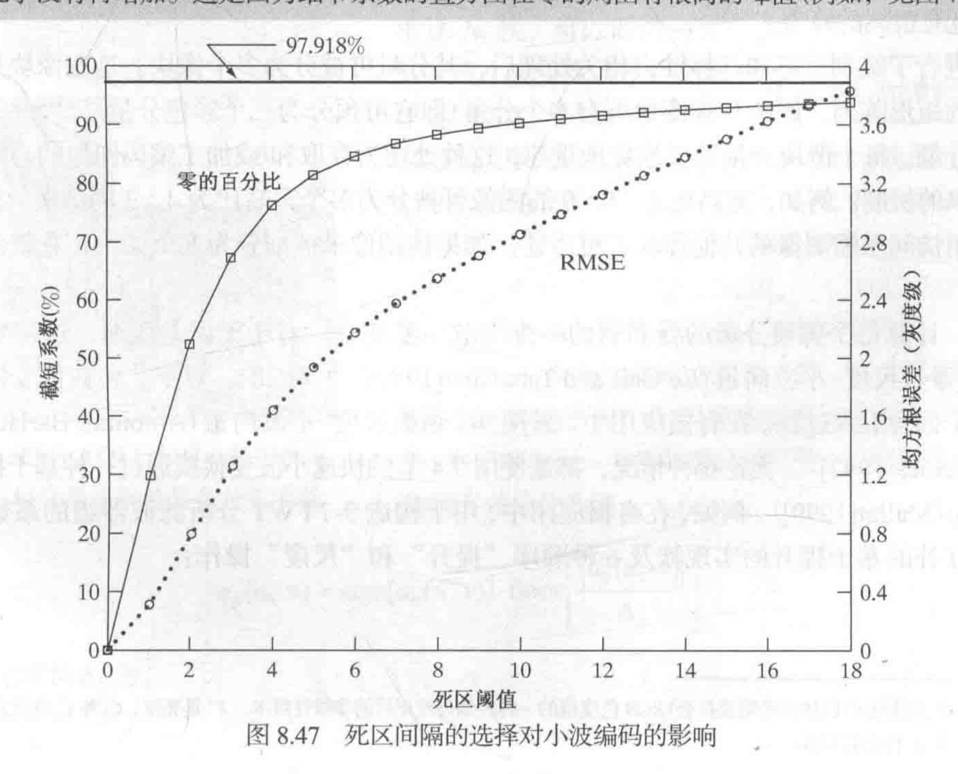


图 8.47　死区间隔的选择对小波编码的影响

628
~
629

图 8.47 中与死区阈值对应的均方根重建误差在阈值 5 处从 0 增大到 1.94 个灰度级，而在阈值 18 处，增大到 3.83 个灰度级，在这一点零的数量达到 93.85%。如果每个细节系数都被消掉，即百分比上升到 97.92%(上升约 4%)，则重建误差会增加到 12.3 个灰度级。

JPEG-2000

JPEG-2000 扩充了流行的 JPEG 标准，在连续色调静止图像的压缩和压缩数据的访问方面提供了更大的灵活性。例如，可以提取 JPEG-2000 压缩图像中的一部分进行转发、存储、显示和/或编辑。该标准以刚刚讨论的小波编码技术为基础。系数量化与单一尺度和子带相适应，并且在比特平面(见 8.2.7 节)的基础上对量化后的系数进行算术编码。使用标准符号，一幅图像可以进行如下编码(ISO/IEC[2000])。

编码过程的第一步是，通过减去 2^{Ssiz-1}，对被编码的 *Ssiz* 比特的无符号图像的样本进行直流电平平移。如果该图像具有多于一个的分量（如彩色图像的红色、绿色和蓝色平面），则单独平移每个分量。如果恰好有三个分量，则可使用一个可逆的或非可逆的线性组合，来对它们有选择地进行去相关处理。例如，标准中非可逆的分量变换是

> 该标准中使用 *Ssiz* 来表示灰度分辨率。

$$
\begin{aligned}
Y_0(x, y) &= 0.299 I_0(x, y) + 0.587 I_1(x, y) + 0.114 I_2(x, y) \\
Y_1(x, y) &= -0.16875 I_0(x, y) - 0.33126 I_1(x, y) + 0.5 I_2(x, y) \\
Y_2(x, y) &= -0.5 I_0(x, y) - 0.41869 I_1(x, y) - 0.08131 I_2(x, y)
\end{aligned}
\tag{8.2-61}
$$

式中，I_0, I_1 和 I_2 是级别平移过的输入分量，Y_0, Y_1 和 Y_2 是相应的去相关分量。如果输入分量是彩色图像的红色、绿色和蓝色平面，则式(8.2-61)用 $Y'C_bC_r$ 的彩色视频变换近似 $R'G'B'$ (Poynton[1996])①。变换的目的是改善压缩效率；变换后的分量 Y_1 和 Y_2 是差值图像，该差值图像的直方图在零点附近有很高的峰值。

> 不可逆分量变换是用于有损压缩的分量变换。该分量本身是可逆的。可逆压缩使用了一种不同的分量变换。

图像进行了级别平移和选择性去相关处理后，其分量可被分为多个像块。这些像块是被单独处理的像素的矩形阵列。因为一幅图像可有多个分量(即它可预分为三个彩色分量)，因此，像块处理生成了块分量。每个像块分量可以独立地重建，这种处理为存取和/或加工编码图像的有限区域提供了一种简单的机制。例如，宽高比为 16：9 的图像可被分为多个宽高比为 4：3 的像块。然后，该像块可以不用访问压缩图像的其他像块就可重建。如果该图像未被细分为多个像块，它就是一个单一的像块。

然后，计算每个像块分量的行和列的一维离散小波变换。对于无误差压缩，这种变换是以双正交、5-3 系数尺度-小波向量(Le Gall and Tabatabai[1988])为基础的。对于非整数值变换系数，还定义了一个四舍五入过程。在有损应用中，采用 9-7 系数尺度-小波向量(Antonini, Barlaud, Mathieu and Daubechies[1992])。无论哪种情况，都是使用 7.4 节的快速小波变换或通过一种基于提升的方法进行计算的(Mallat[1999])。例如，在有损应用中，用于构造 9-7 FWT 分析滤波器组的系数在表 8.15 中给出。互补的基于提升的实现涉及 6 种顺序“提升”和“尺度”操作：

629 ~ 631

① $R'G'B'$ 是线性 CIE(国际照明委员会)*RGB* 色度值的一种经伽马校正后的非线性版本。Y' 是亮度，C_b 和 C_r 是色差(即 $B'-Y'$ 和 $R'-Y'$ 经标定后的值)。

$$
\begin{aligned}
&Y(2n+1) = X(2n+1) + \alpha\left[X(2n) + X(2n+2)\right], && i_0 - 3 \leqslant 2n+1 < i_1 + 3 \\
&Y(2n) = X(2n) + \beta\left[Y(2n-1) + Y(2n+1)\right], && i_0 - 2 \leqslant 2n < i_1 + 2 \\
&Y(2n+1) = Y(2n+1) + \gamma\left[Y(2n) + Y(2n+2)\right], && i_0 - 1 \leqslant 2n+1 < i_1 + 1 \\
&Y(2n) = Y(2n) + \delta\left[Y(2n-1) + Y(2n+1)\right], && i_0 \leqslant 2n < i_1 \\
&Y(2n+1) = -K \cdot Y(2n+1), && i_0 \leqslant 2n+1 < i_1 \\
&Y(2n) = Y(2n)/K, && i_0 \leqslant 2n < i_1
\end{aligned}
\tag{8.2-62}
$$

表 8.15 不可逆 9-7 小波变换的低通和高通分析滤波器的冲激响应

滤波器抽头	高通小波系数	低通尺度系数
0	−1.115 087 052 456 994	0.602 949 018 236 357 9
±1	0.591 271 763 114 247 0	0.266 864 118 442 872 3
±2	0.057 543 526 228 499 57	−0.078 223 266 528 987 85
±3	−0.091 271 763 114 249 48	−0.016 864 118 442 874 95
±4	0	0.026 748 757 410 809 76

式中，X 是变换的像块分量，Y 是变换结果，i_0 和 i_1 定义了分量内的像块分量的位置。也就是说，它们是变换像块分量的行或列的第一个取样及紧接着的下一个取样的索引。变量 n 基于 i_0，i_1 假定和执行的 6 个操作值。如果 $n< i_0$ 或 $n \geqslant i_1$，则 $X(n)$ 通过对称地扩充 X 来得到。例如，$X(i_0-1)=X(i_0+1)$，$X(i_0-2)=X(i_0+2)$，$X(i_1)=X(i_1-2)$，$X(i_1+1)=X(i_1-3)$。提升和尺度操作的结果，Y 的偶索引值等于 FWT 低通滤波后的输出；而 Y 的奇索引值对应于高通 FWT 滤波后的结果。提升参数 α, β, γ 和 δ 分别为−1.586 134 342, −0.052 980 118, 0.882 911 075 和 0.433 506 852。尺度因子 K 为 1.230 174 105。

这些基于提升的系数是在标准中指定的。

刚才描述的变换生成 4 个子带，分别是像块分量的低分辨率近似及分量的水平、垂直和对角频率特征。将后续迭代限制为前面的分解的近似系数，把该变换重复 N_L 次，可以得到一个 N_L 尺度小波变换。相邻尺度以 2 的幂次空间相联系，并且最低尺度仅包含显式定义的原始像块分量的近似。如在图 8.48 中推测的那样，这里总结了 $N_L=2$ 时的 JPEG-2000 标准的符号。一个一般的 N_L 尺度变换包含 $3N_L+1$ 个子带，这些子带的系数对于 $b = N_LLL, N_LHL, \cdots, 1HL, 1LH, 1HH$ 表示为 a_b。该标准没有规定需要计算的尺度数量。

回顾第 7 章可知，DWT 将一幅图像分解为一组称为子带的带限分量。

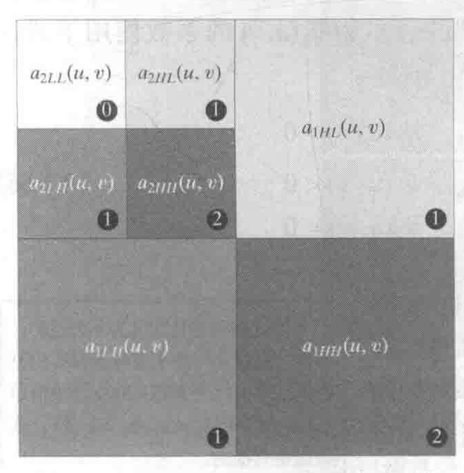


图 8.48 JPEG-2000 二尺度小波变换像块分量系数符号和分析增益

当每个块分量都经过处理之后，变换系数的总数等于原图像中的取样数——但重要的视觉信息集中于少数系数中。为了减少表示变换所需的比特数，子带 b 的系数 $a_b(u,v)$ 使用下式量化为值 $q_b=(u,v)$：

$$
q_b(u,\ v) = \mathrm{sign}[a_b(u,\ v)] \cdot \mathrm{floor}\left[\frac{|a_b(u,\ v)|}{\Delta_b}\right] \tag{8.2-63}
$$

式中，量化步长 Δ_b 为

631 ~ 632

$$\Delta_b = 2^{R_b - g_b}\left(1 + \frac{\mu_b}{2^{11}}\right) \tag{8.2-64}$$

R_b 是子带 b 的标定动态范围，而 ε_b 和 μ_b 是分配给子带系数的指数和尾数的比特数。子带 b 的标定动态范围是用于表示原图像的比特数和子带 b 的分析增益比特数之和。子带分析增益比特遵从图 8.48 所示的简单模式。例如，子带 $b = 1HH$ 有两个分析增益比特。

不要混淆标准中对标称动态范围的定义与第 2 章中紧密相关的定义。

对无误差压缩，有 $\mu_b = 0$, $R_b = \varepsilon_b$ 和 $\Delta_b = 1$。对不可逆压缩，标准中没有指定特定的量化步长。相反，必须以子带为基础向解码器提供指数和尾数的比特数，这称为解释式量化，或仅针对 N_LLL 子带，这称为导出式量化。在后一种情况下，对余下的子带使用外推的 N_LLL 子带参数进行量化。令 ε_0 和 μ_0 表示分配给 N_LLL 子带的比特数，则子带 b 的外推参数为

$$\begin{aligned} \mu_b &= \mu_0 \\ \varepsilon_b &= \varepsilon_0 + n_b - N_L \end{aligned} \tag{8.2-65}$$

式中，n_b 表示从原图像像块分量到子带 b 的子带分解级别的数量。

在编码过程的最后一步中，每个变换像块分量的子带的系数被排列为称为码块的矩形块，这些码块被单独编码，一次一个比特平面。从带有非零元素的最高阶比特平面开始，每个比特平面都在 3 个通道中处理。(比特平面中的)每个比特仅在 3 个通道之一中被编码，这 3 个通道分别称为有效性传播、幅度细化和提纯。然后，对输出进行算术编码，并对其他码块使用相似的途径进行组合以形成层。层是来自每个代码块的编码通道的任意数量的组合。得到的层最终被分割成包，从而提供从总码流中提取感兴趣空间区域的附加方法。

JPEG2000 解码器是前述过程的简单反操作。重建来自算术编码的 JPEG2000 包的块分量子带后，用户选择的一些子带来解码。虽然对于特殊子带编码器可能已编码 M_b 个比特平面，但由于码流的嵌入性质，用户可能选择只解码 N_b 个比特平面，这相当于使用 $2^{M_b - N_b} \cdot \Delta_b$ 的步长来量化码块系数。所有未解码的比特都置为零，且所得到的表示为 $\overline{q}_b(u,v)$ 的系数使用下式进行反量化：

$$R_{q_b}(u,\, v) = \begin{cases} (\overline{q}_b(u,\, v) + r \cdot 2^{M_b - N_{b(u,v)}}) \cdot \Delta_b, & \overline{q}_b(u,\, v) > 0 \\ (\overline{q}_b(u,\, v) - r \cdot 2^{M_b - N_{b(u,v)}}) \cdot \Delta_b, & \overline{q}_b(u,\, v) < 0 \\ 0, & \overline{q}_b(u,\, v) = 0 \end{cases} \tag{8.2-66}$$

式中，$R_{q_b}(u,\, v)$ 表示一个反量化的变换系数，而 $N_b(u,\, v)$ 是 $\overline{q}_b(u,v)$ 的已解码比特平面数。重建参数 r 通过产生最好视觉或客观重建质量的解码器来选择。通常，$0 \leqslant r < 1$，常用值为 1/2。然后，反量化系数使用一个 FWT^{-1} 滤波器组按列和按行进行逆向变换，这些系数根据表 8.15 和式(7.1-11)或通过下列基于提升的操作得到：

本章之前将量化定义为不可逆的。术语“反量化”并不意味着没有信息损失。除不可逆 JPEG-2000 压缩情形(这时 $\mu_b = 0, R_b = \varepsilon_b, \Delta_b = 1$)外，该处理是有损的。

$$\begin{aligned} &X(2n) = K \cdot Y(2n), && i_0 - 3 \leqslant 2n < i_1 + 3 \\ &X(2n+1) = (-1/K) \cdot Y(2n+1), && i_0 - 2 \leqslant 2n - 1 < i_1 + 2 \\ &X(2n) = X(2n) - \delta[X(2n-1) + X(2n+1)], && i_0 - 3 \leqslant 2n < i_1 + 3 \\ &X(2n+1) = X(2n+1) - \gamma[X(2n) + X(2n+2)], && i_0 - 2 \leqslant 2n + 1 < i_1 + 2 \\ &X(2n) = X(2n) - \beta[X(2n-1) + X(2n+1)], && i_0 - 1 \leqslant 2n < i_1 + 1 \\ &X(2n+1) = X(2n+1) - \alpha[X(2n) + X(2n+2)], && i_0 \leqslant 2n + 1 < i_1 \end{aligned} \tag{8.2-67}$$

633
~
634

式中，参数 $\alpha, \beta, \gamma, \delta$ 和 K 的定义与式(8.2-62)中的相同。反量化后的系数行或列元素 $Y(n)$在需要时可以对称扩展。解码的最后一步是分量像块的组合，反向分量变换(如果需要的话)和直流(DC)水平平移。对于不可逆编码，分量反变换是

$$
\begin{aligned}
I_0(x, y) &= Y_0(x, y) + 1.402Y_2(x, y) \\
I_1(x, y) &= Y_0(x, y) - 0.34413Y_1(x, y) - 0.71414Y_2(x, y) \\
I_2(x, y) &= Y_0(x, y) + 1.772Y_1(x, y)
\end{aligned}
\tag{8.2-68}
$$

变换后的像素被平移了 $+2^{Ssiz-1}$。

例 8.28　基于小波的 JPEG-2000 编码与基于 DCT 的 JPEG 压缩的比较。

图 8.49 显示了图 8.9(a)中单色图像的 4 幅 JPEG-2000 近似。图中相继各行说明压缩水平递增——包括 $C = 25, 52, 75$ 和 105。第一列图像是 JPEG-2000 编码解压缩后的结果。这些图像与原图像[见图 8.9(a)]的差示于第二列，第三列包含了第一列重建图像的局部放大部分。因为前两行的压缩率与例 8.18 中的压缩率相同，这些结果可以与图 8.32(a)到(f)中基于 JPEG 变换的结果进行定性和定量比较。

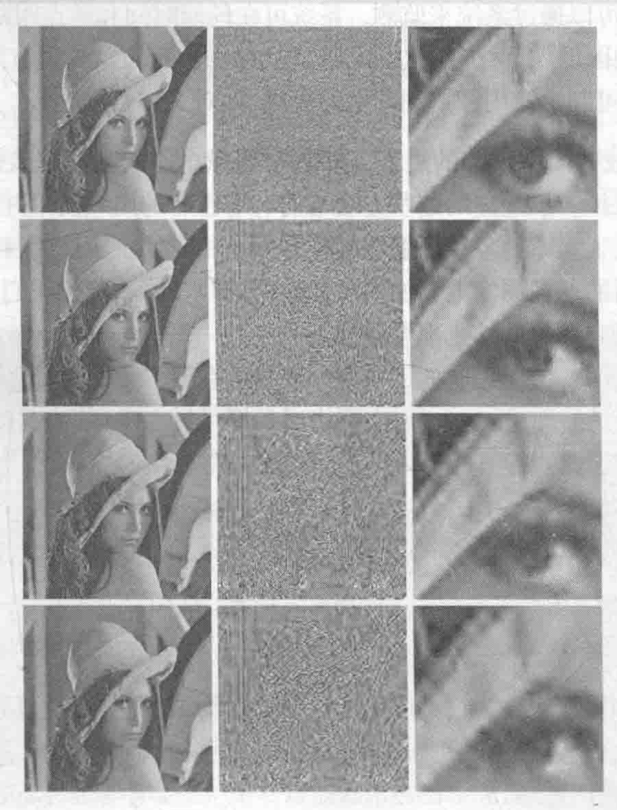


图 8.49　图 8.9(a)的 4 幅 JPEG-2000 近似。每一行均包含了压缩和重建后的结果、该结果与原图像间经过缩放的差值图像，以及重建图像的局部放大部分(请将第一行和第二行与图 8.32 中的 JPEG 结果进行比较)

图 8.49 第一行和第二行中的误差图像与图 8.32(b)和(e)中对应图像的视觉比较表明，JPEG-2000 结果中的误差明显减小——JPEG 结果的误差分别为 5.4 个灰度级和 10.7 个灰度级，而 JPEG-2000 的误差分别为 3.86 个灰度级和 5.77 个灰度级。在这两种压缩水平上，计算的误差有利于基于小波的结果。除了降低重建误差外，小波编码引人注目地增强了图像的质量(在主观感觉上)。注意，JPEG 结果中明显的块效应[见图 8.32(c)和(f)]在图 8.49 中并不存在。最后，我们注意到，图 8.49 中第三行和第四行所能达到的压缩用 JPEG 并不实际。JPEG-2000 可提供压缩率超过 100 : 1 的可用图像——最讨厌的降质就是图像模糊的增加。

634 ~ 636

8.3 数字图像水印

8.2 节中的方法和标准构成了图像(图片或视频)在数字媒体和互联网上的实际发布。遗憾的是，这样发布的图像可被重复且无误差地复制，这就把它们的所有者的权利置入风险之中。甚至为发布加密后，解密后的图像也不能得到保护。阻止非法复制的方法之一是，把称为水印的一条或多条信息项插入到易受攻击的图像中，在这种方法中，水印不能从图像本身分离出来。如同水印图像的整体部分那样，它们以各种方法来保护所有者的权益，包括：

1. 版权识别。当所有者的权益受到侵犯时，数字水印可提供作为所有权证据的信息。
2. 用户识别或指纹。合法用户的身份可以在水印中编码，并用于识别非法复制源。
3. 著作权认定。水印的存在可保证一幅图像不被篡改——假定水印被设计成对图像的任何修改都将破坏水印。
4. 自动监视。水印可以通过系统来监测，系统可在任何时间和地点跟踪所使用的图像(可搜索放在 Web 网页上的图像)。对于征收版税和/或非法用户定位监测很有用。
5. 复制保护。水印可制定使用和复制图像的规则(如对 DVD 播放器)。

本节简要介绍提供数字图像水印的过程，即数字图像水印处理是把数据插入到一幅图像中的过程，进而形成对图像的保护。描述过的这些方法与前一节已有的压缩技术并无相同之处，但它们的确都涉及信息编码。事实上，水印处理和压缩在某些方面是相对的。压缩的目的是减少用于表示图像的数据量，但水印处理的目的是在图像上添加信息和数据(即水印)。正如我们在本节其余内容中所看到的那样，水印本身可以是完全可见的或不可见的。

可见水印是一幅不透明的或半透明的子图像，或是放在另一幅图像(即用水印处理的图像)之上的图像，因此，它对观察者是显而易见的。电视网络通常把可见水印(依照其标识风格)放在屏幕的右上角或右下角。如下面的例子表明的那样，可见水印处理通常在空间域执行。

例 8.29 简单的可见水印。

图 8.50(b)中的图像是在图8.9(a)中图像的右下四分之一处叠放了图8.50(a)中水印的缩小版本。令 f_w 代表水印的图像，我们可把它表示为未加水印图像 f 和水印 w 的如下线性组合：

$$f_w = (1-\alpha)f + \alpha w \tag{8.3-1}$$

式中，常数 α 控制水印和衬底图像的相对可见性。如果 α 为 1，则水印是不透明的，并且衬底图像完全是暗的。随着 α 接近 0，会逐渐看到更多的衬底图像和更少的水印。通常，$0<\alpha \leqslant 1$；在图 8.50(b)中，$\alpha = 0.3$。图 8.50(c)是图 8.50(b)中已加水印图像和图 8.9(a)中未加水印图像的差。灰度 128 表示 0 差值。注意，衬底图像透过半透明水印清晰可见。这在图 8.50(b)中和图 8.50(c)中所示的差值图像中是很明显的。

与前面例子的可见水印不同，不可见水印裸眼是看不见的。它们不能被感知，但使用合适的解码算法可恢复它。不可见性是通过作为可见的冗余信息插入水印来保证的，如人眼视觉系统所忽略的或不能感知(见 8.1.3 节)的信息。图 8.51(a)提供了一个简单的例子。因为 8 比特图像中的最低阶比特对我们感知图像实质上没有效果，来自图 8.50(a)的水印被插入或“隐藏”在两个最低阶比特中。使用上面介绍的符号，我们令

636 ~ 638

$$f_w = 4\left(\frac{f}{4}\right) + \frac{w}{64} \tag{8.3-2}$$

并使用无符号整数算术来执行计算。用4除f，再乘4，置f的两个最低阶比特为0，用64除w，把它的两个最高阶比特移到两个最低阶比特的位置，再把两个结果相加，产生LSB水印图像。注意，嵌入的水印在图8.51(a)中不可见。把这一图像的最高有效6比特置0，并把其余的值放大到整个灰度范围，就可提取出水印，如图8.51(b)所示。



图8.50 一个简单的可见水印：(a)水印；(b)加过水印的图像；(c)加过水印的图像和原图像(未加水印图像)间的差

不可见水印的一个重要特性是，它能抵抗偶然和故意删除它们的企图。对嵌入了易碎的、不可见水印的图像的任何修改，都会破坏这种水印。在某些应用中，如图像鉴定，这是所期望的一种特性。如图8.51(c)和(d)所示，图8.51(a)中LSB嵌入过水印的图像包含易碎的不可见水印。如果图8.51(a)中的图像用有损JPEG来压缩和解压缩，就会破坏水印。图8.51(c)是图8.51(a)经压缩和解压缩后的结果；均方根误差(rms)是2.1比特。如果我们试图用与图8.51(b)中相同的方法从这幅图像中提取水印，那么结果是不可理解的[见图8.51(d)]。虽然有损压缩和解压缩保留了图像中的重要视觉信息，但会破坏易碎的水印。

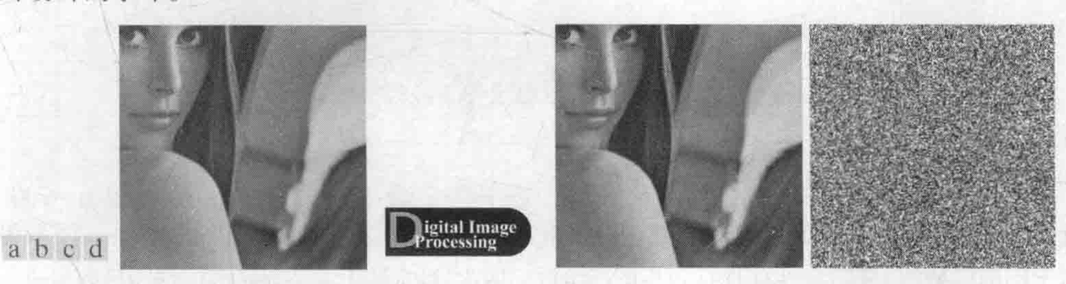


图8.51 一个简单的不可见水印：(a)加过水印的图像；(b)提取的水印；(c)高质量JPEG压缩和解压缩后加过水印的图像；(d)从图(c)中提取的水印

鲁棒的不可见水印设计用于使图像能经受修改，而不管所谓的攻击是有意的还是无意的。常见的无意攻击包括有损压缩、线性和非线性滤波、修剪、旋转、重取样和类似的操作。故意攻击包括从打印和重扫描到添加附加的水印和/或噪声等。当然，使图像本身处在不可用状态的抗攻击是没有必要的。

图8.52显示了一个典型图像水印处理系统的基本组成部分。图8.52(a)中的编码器把水印w_i插入到图像f_i中，产生加过水印的图像f_{w_i}；图8.52(b)中的解码器提取和验证加过水印的输入f_{w_i}或没有加过水印的输入f_i中w_i的存在性。如果w_i可见，则不需要解码器；如果不可见，解码器可能需要也可能不需要f_i和w_i的副本[在图8.52(b)中以灰色显示]去做提取或验证工作。如果使用了f_i和/或w_i，则水印处理系统是已知的私钥系统或受限密钥系统；如果不用，则称为公钥系统或不受限密钥系统。因为解码器必须处理加过水印的图像和没有加过水印的图像，在图8.52中使用$w_\varnothing$来表示不存在水印。最后，我们注意到，为确定图像中是否存在w_i，解码器必须对提取的水印w_j与w_i做相关操作，并将相关操作的结果与一个预定义的阈值进行比较。阈值用于设置一个可接受"匹配"的相似度。

638~639

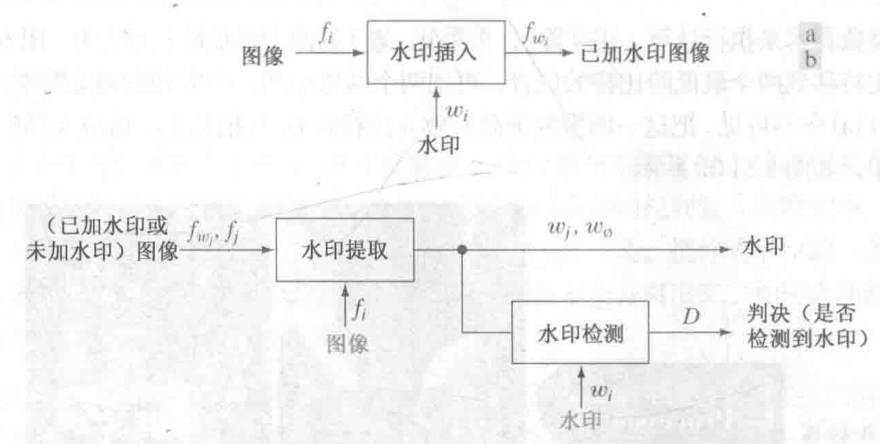


图 8.52 一个典型的图像水印处理系统：(a)编码器；(b)解码器

例 8.30 基于 DCT 的不可见的鲁棒水印。

水印嵌入和提取可如前几个例子那样在空间域执行，或变换域中执行。图 8.53(a)和(c)显示了图 8.9(a)中图像的两个基于 DCT 的水印方法的已加水印版本，方案如下(Cox et al.[1997])：

> 伪随机数序列的性质近似随机数的性质。它并不是真正随机的，因为它取决于一个预先求出的初始值。

步骤 1 计算将被加入水印图像的二维 DCT。

步骤 2 按幅值定位它的 K 个最大系数 $c_1, c_2, \cdots, c_K$。

步骤 3 通过生成一个 K 元素伪随机数序列 $\omega_1, \omega_2, \cdots, \omega_k$ 创建一个水印，这些伪随机数取自均值 $\mu = 0$、方差 $\sigma^2 = 1$ 的一个高斯分布。

步骤 4 使用下式，将步骤 3 得到的水印嵌入到步骤 2 得到的 K 个最大 DCT 系数中：

$$c_i' = c_i \cdot (1 + \alpha\omega_i), \quad 1 \leqslant i \leqslant K \tag{8.3-3}$$

式中，指定常数 $\alpha > 0$（它控制 ω_i 更改 c_i 的范围）。使用式(8.3-3)计算的 c_i' 代替原来的 c_i。

步骤 5 计算来自步骤 4 的结果的反 DFT。

通过采用由伪随机数形成的水印，并把它们散布到一幅图像感知上的重要频率成分上，α 可取得较小，以降低水印的可见性。同时，水印保持较高的安全性，因为(1)水印由无明显结构的伪随机数组成，(2)水印被嵌入到影响整个二维空间图像的多个频率成分中(因此，它们的位置并不明显)，(3)对它们的攻击也会使图像降质(也就是说，图像的大多数重要频率万分必须被改变，以影响水印)。

由于在图 8.53(a)和(c)中图像的 DCT 系数上嵌入了来自伪随机数的水印图像，因此图 8.53(b)和(d)中图像的灰度产生了变化。很明显，伪随机数必须有一个效果，即在加过水印的图像上它很小，以至于看不见。为了显示这一效果，从图 8.9(a)所示没有加过水印的图像中减去图 8.53(a)和(c)中的图像，并将灰度范围放大到区间[0, 255]。图 8.53(b)和(d)是结果图像；它们显示了伪随机数的二维空间贡献。然而，因为它们已被放大，我们不能简单地把这些图像加到图 8.9(a)的图像上，并得到图 8.53(a)和(c)中的水印图像。正如在图 8.53(a)和(c)中看到的那样，它们的实际灰度干扰小到可以忽略不计。

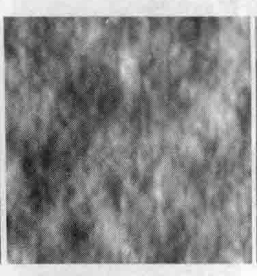
为了确定一幅特定的图像是否是一幅事先加过水印图像的副本，该图像具有水印 $\omega_1, \omega_2, \cdots, \omega_k$ 和 DCT 系数 $c_1, c_2, \cdots, c_K$，我们采用下列步骤：

步骤 1 计算该问题中图像的二维 DCT。

步骤 2 提取 K 个 DCT 系数(对应于水印处理过程中步骤 2 的 $c_1, c_2, \cdots, c_K$ 的位置)，并将这些系数表示为 $\hat{c}_1, \hat{c}_2 \cdots, \hat{c}_K$。如果问题图像是事先加过水印的图像(未被修改)，则对于 $1 \leqslant i \leqslant K$ 有 $\hat{c}_i = c_i'$。如果它是加过水印图像的一个修改过的副本(即它受到过某种攻击)，则对于 $1 \leqslant i \leqslant K$ 有 $\hat{c}_i \approx c_i'$（$\hat{c}_i$ 是 c_i'

639 ~ 641

的近似)。否则，问题图像将是一幅无水印图像，或是一幅具有完全不同水印的图像——$\hat{c}_i$ 将不会类似于原始的 c_i'。



a b c d

图 8.53　(a)和(c)图 8.9(a)的两个加过水印的版本；(b)和(d)加过水印的图像和没有加过水印的图像间的差(灰度被放大了)。这两幅图像显示了伪随机水印对原图像的灰度贡献(尽管缩放显著)

步骤 3　使用下式计算水印 $\hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_K$：

$$\hat{\omega}_i = \frac{\hat{c}_i - c_i}{\alpha c_i}, \quad 1 \leqslant i \leqslant K \tag{8.3-4}$$

回顾可知，水印是一个伪随机数序列。

步骤 4　用一个参数，如相关系数，度量 $\hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_K$ (来自步骤 2)与 $\omega_1, \omega_2, \cdots, \omega_K$ (来自嵌入水印过程的步骤 3)的相似度，

12.2.1 节中将详细探讨相关系数。

$$\gamma = \frac{\sum_{i=1}^{K}(\hat{\omega}_i - \overline{\hat{\omega}})(\omega_i - \overline{\omega})}{\sqrt{\sum_{i=1}^{K}(\omega_i - \overline{\hat{\omega}})^2 \cdot \sum_{i=1}^{K}(\omega_i - \overline{\omega})^2}}, \quad 1 \leqslant i \leqslant K \tag{8.3-5}$$

式中，$\overline{\omega}$ 和 $\overline{\hat{\omega}}$ 是两个 K 元素水印的均值。

步骤 5　将度量的相似度 γ 与一个预定义的阈值 T 进行比较，并进行二值检测判决：

$$D = \begin{cases} 1, & \gamma \geqslant T \\ 0, & \text{其他} \end{cases} \tag{8.3-6}$$

换句话说，$D = 1$ 表明水印 $\omega_1, \omega_2, \cdots, \omega_K$ 存在(关于特定的阈值 T)；$D = 0$ 则表明水印不存在。

使用这一过程，图 8.53(a)中加过水印的原始图像，针对其本身度量，得到一个相关系数 0.999 9，即 $\gamma = 0.9999$。它是一个无错误的匹配。采用相似的方法，当针对图8.53(a)中的图像度量时，图8.53(b)中的图像会得到 $\gamma = 0.0417$，对于图 8.53(a)中加过水印的图像，它不可能被弄错，因为相关系数很小。

在结束本节时，我们注意到，前例中的基于 DCT 的水印处理方法可完全抵抗水印攻击，部分原因是，这是一种私钥方法或受限密钥方法。受限密钥方法总是要比非受限密钥方法灵活。使用图 8.53(a)中的加过水印的图像，图 8.54 说明了该方法抵抗各种常见攻击的能力。如图中所看到的那样，水印检测在被执行的攻击范围之外要好得多——得到的相关系数(显示在图中每幅图像的下方)在 0.3113 到 0.9945 间变化。在受到高质量但有损(导致 7 个灰度级的均方根误差)的 JPEG 压缩和解压缩时，$\gamma = 0.9945$。甚至当压缩和重建产生 10 个灰度级的均方根误差时，$\gamma = 0.7395$——该图像的可用性明显降低。使用空间滤波的有效平滑和加入高斯噪声，都未使相关系数降到 0.8230 以下。然而，直方图均衡会把 γ 减小到 0.5210；而旋转的影响最大，它把 γ 减小到 0.3313。除了图 8.53(a)中的有损 JPEG 压缩和重建外，所有攻击都大大降低了原始加过水印的图像的可用性。

641 ~ 642



图 8.54 对图 8.53(a)中加过水印的图像的攻击：(a)具有 7 个灰度级均方根误差的有损 JPEG 压缩和解压缩图像；(b)具有 10 个灰度级的均方根误差的有损 JPEG 压缩和解压缩图像(注意块效应)；(c)用空间滤波方法平滑处理的结果；(d)加入高斯噪声的图像；(e)直方图均衡后的图像；(f)旋转后的图像。每一幅图像都是图 8.53(a)中加过水印图像的一个修改版本。修改之后，它们不同程度地保留了各自的水印，如每幅图像下方的相关系数指出的那样

小结

本章的主要目的是介绍数字图像压缩的理论基础，描述最常用的压缩方法，并介绍数字图像水印处理的相关领域。尽管论述是介绍性的，但参考资料提供了这一主题的广泛性论述。如表 8.3 和表 8.4 中所列国际标准表明的那样，压缩方法在文献图像存储与传送、互联网和商业视频传播(如 DVD)方面有着重要作用。压缩也是少数几个在商业上广泛要求对普遍接受的标准进行授权使用的图像处理领域之一。由于越来越多的图像以压缩的数字格式进行传播，因此图像水印处理变得越来越重要。

参考文献

本章 8.1 节的介绍性材料是图像压缩的基础，第 1 章末尾给出的多数图像处理书籍中，都能找到不同形式的内容。关于人类视觉系统的其他信息，见 Netravali and Limb[1980]、Huang[1966]、Schreiber and Knapp[1958]及第 2 章结尾处给出的文献。关于信息论的更多内容，见本书的网站或 Abramson[1963]、Blahut[1987]和 Berger[1971]。香农奠定该领域基础的的经典论文“通信中的数学理论”(*A mathematical Theory of Communication*)[1948]，也是另一份优秀的参考资料。主观保真度准则详见 Frendenall and Behrend[1960]。

贯穿全章的许多例子中使用了各种压缩标准。大多数例子都是用 Adobe Photoshop(有能免费使用的压缩插件)和/或 MATLAB 实现的，详见 Gonzalez et al.[2004]。一般来说，压缩标准是冗长且复杂的；我们不可能全部覆盖它们。关于某个特殊标准的详细信息，可参阅相关标准组织（国际标准化组织、国际电子技术委员会和/或国际电信联盟）发布的文件。标准方面的其他参考文献有 Hunter and Robinson[1980]、Ang et al.[1991]、Fox[1991]、Pennebaker and Mitchell[1992]、Bhatt et al.[1997]、Sikora [1997]、Bhaskaran and Konstantinos[1997]、Ngan et al.[1999]、Weinberger et al.[2000]、Symes[2001]、Mitchell et al.[1997]和 Manjunath et al.[2001]。

643 ~ 644

8.2 节中描述的有损和无误差编码技术与 8.3 节中的水印处理技术，大部分源于本书引用的原始论文。涵盖的算法是这一领域中具有代表性的研究成果，但不是全部成果。关于 LZW 编码的素材，源自 Ziv and Lempel[1977, 1978]。算术编码的内容源自 Witten, Neal and Cleary[1987]。算术编码的一种更为重要的实现见 Pennebaker et al.[1988]。对于无损预测编码的较好论述，见 Rabbani and Jones[1991]。式(8.2-56)的自适应预测器来自 Graham[1958]。关于运动补偿的更多内容，见 S. Solari[1997](它还包含普通视频压缩和压缩标准的介绍)和 Mitchell et al.[1997]。8.3 节中基于 DCT 的水印处理技术源于 Cox et al.[1997]。关于水印处理的更多内容，见 Cox et al.[2001]、Parhi and Nishitani[1999]和 S. Mohanty[1999]。

图像压缩领域的综述性文章也有很多，包括 Netravali and Limb[1980]、A. K. Jain[1981]、*IEEE Transactions on Communications*[1981]图像通信系统专刊、*Proceedings of IEEE*[1980]图形编码专刊、*Proceedings of the IEEE*[1985]视觉通信系统专刊、*IEEE Transactions on Image Processing* [1994]图像序列压缩专刊，以及 *IEEE Transactions on Image Processing*[1996]矢量量化专刊。此外，*IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Viedo Technology* 和 *IEEE Transactions on Multimedia* 等刊物中也包含有关于视频和静止图像压缩、运动补偿和水印处理方面的文章。例如，Robinson[2006]、Chandler and Hemami[2005]、Yan and Cosman [2003]、Boulgouris et al.[2001]、Martin and Bell [2001]、Chen and Wilson[2000]、Hartenstein et al.[2000]、Yang and Ramchandran [2000]、Meyer et al.[2000]、S. Mitra et al. [1998]、Mukherjee and Mitra[2003]、Xu et al.[2005]、Rane and Sapiro[2001]、Hu et al.[2006]、Pi et al.[2006]、Dugelay et al.[2006]和 Kamstra and Heijmans[2005]，都可作为深入阅读和参考的起点。

习题

8.1 (a) 变长编码过程能否用于压缩具有 2^n 个灰度级直方图均衡过的图像？请解释其原因。

(b) 这样一幅图像是否包含可以用于数据压缩的空间和时间冗余？

8.2 行程编码的一种变形是：(1) 只对 0 行程或 1 行程(而非两者)编码，(2) 在每行的开始处分配一个特殊的码，以减小传输误差的影响。一种可能的编码对是 (x_k, r_k)，其中 x_k 和 r_k 分别表示第 k 个行程的起始坐标和行程。编码(0, 0)用于表示每个新行。

(a) 当对一幅大小为 $2^n \times 2^n$ 的二值图像进行行程编码时，推导为确保数据压缩而要求的每条扫描线最大平均行程的一般表达式。

(b) 计算 $n=8$ 时的最大允许值。

8.3 考虑灰度数据为{255, 118, 127, 182, 18, 178, 82, 55}的一条 8 像素的线。如果使用 4 比特精度来均匀量化它，计算量化后的数据的均方根误差和均方根信噪比。

★8.4 虽然量化会导致信息损失，但对人眼来说有时它是不可见的。例如，当 8 比特像素被均匀量化为更少的比特/像素时，通常会出伪轮廓。使用改进的灰度缩放(IGS)量化方法，可以减少或消除它。一个和值(开始时置 0)是由当前的 8 比特灰度值和前一个生成的和的 4 个最低阶比特形成的。如果灰度值的 4 个最高阶比特是 1111_2，因此加 0000_2 替换它。产生的和的 4 个最高阶比特被用做编码后的像素值。

(a) 对灰度数据{108, 139, 135, 244, 172, 178, 56, 97}构造 IGS 编码。

(b) 对解码后的 IGS 数据，计算均方根误差和均方根信噪比。

8.5 对大小为 1024×1024、熵为 4.2 比特/像素［使用式(8.1-7)从其直方图计算得到］的 8 比特图像进行霍夫曼编码。

(a) 所期望的最大压缩率是多少？

(b) 它能得到吗？

(c) 如果要求较高级别的无损压缩，还应该做什么？

644
~
645

★8.6 以 e 为底的信息单位通常称为奈特(nat)，以 10 为底的信息单位则称为哈特利(Hartley)。计算将这些单位与以 2 为底的信息单位(比特)联系起来的转换因子。

★8.7 证明，对于有 q 个符号的零记忆信源，熵的最大值为 $\log q$。当且仅当所有信源符号等概率出现时，熵才能达到最大值。[提示：考虑量 $\log q - H(z)$并注意不等式 $\ln x \leqslant x-1$。]

8.8 (a) 对于一个 4 符号信源，有多少唯一的霍夫曼编码?

(b) 构建它们。

8.9 考虑如下大小为 4×8 的 8 比特图像：

21	21	95	95	169	169	243	243
21	21	95	95	169	169	243	243
21	21	95	95	169	169	243	243
21	21	95	95	169	169	243	243
21	21	95	95	169	169	243	243

(a) 计算该图像的熵。

(b) 用霍夫曼编码压缩该图像。

(c) 计算霍夫曼编码能达到的压缩率和效率。

★(d) 考虑对一对像素而非单个像素进行霍夫曼编码。也就是说，考虑由产生原图像的零记忆信源的第二次扩展产生的图像。当作为像素对看待时，图像的熵是什么?

(e) 考虑对相邻像素间的差值进行编码。新的差值图像的熵是什么? 关于压缩图像，它告诉我们做什么?

(f) 解释(a)，(d)和(e)中熵的差别。

8.10 用图 8.8 中的霍夫曼码，对编码串 01010010000011101011 进行解码。

8.11 对于 $0 \leqslant n \leqslant 20$，计算 Golomb 码 $G_5(n)$

8.12 针对 Golomb 码 $G_m(n)$，写出解码的通用过程。

8.13 为什么用式(8.2-2)的几何概率密度函数不能计算 $n \geqslant 0$ 的非负整数的霍夫曼码?

8.14 对于 $0 \leqslant n \leqslant 15$，计算指数 Golomb 码 $G_{\exp}^{1}(n)$。

★8.15 针对指数 Golomb 码 $G_{\exp}^{k}(n)$，写出解码的通用过程。

8.16 在式(8.2-3)中，对于 $0<\rho<1$，画出作为 ρ 的函数的最佳 Golomb 编码参数 m 的曲线。

8.17 给定一个 4 符号信源{a, b, c, d}，信源概率为{0.1, 0.4, 0.3, 0.2}，对序列 *abcda* 进行算术编码。

★8.18 算术解码过程是与编码过程相反的过程。用给定的编码模型，对消息 0.32256 进行解码。

符 号	概 率
a	0.2
e	0.3
i	0.1
o	0.2
u	0.1
!	0.1

8.19 使用 8.2.4 节中的 LZW 编码算法，对 7 比特 ASCII 码字符串“AAAAAAAAAAA”进行编码。

★8.20 为例 8.7 的 LZW 编码输出的解码操作设计一种算法。由于编码期间使用的字典此时不能使用，所以在对输出进行解码时必须重新生成码书。

8.21 对 BMP 编码序列{127, 0, 5, 25, 29, 40, 103, 52, 75, 82}进行解码。

8.22 (a) 构造完整的 4 比特格雷码。

(b) 创建一个通用的过程，将格雷码的数字转换成与其等价的二进制码，并用它对 0111010100111 进行解码。

645 ~ 647

8.23 使用 CCITT Group 4 压缩算法对如下两行的第二行进行编码：

0110011100111111100001

1111100011100011111000

假设初始参考元素 a_0 位于第二行的第一个元素上。

8.24 (a) 列出 JPEG DC 系数第 3 类差的所有成员。

(b) 使用表 A.4 计算它们的默认霍夫曼编码。

8.25 使用 MAD 最佳准则、单像素精度和最大允许的 16 个像素的位移，找到一个 8×8 宏块的最佳运动向量需要多少计算？对于 1/4 像素精度，它变成多少次计算？

8.26 对于运动补偿，使用 B 帧的优点是什么？

8.27 画出针对图 8.39 中编码器的伴随运动补偿视频解码器的方框图。

8.28 用一个二阶预测器对自相关函数是式(8.2-49)在 $\rho_h = 0$ 时的形式的一幅图像进行 DPCM 编码。

(a) 构造自相关矩阵 $\boldsymbol{R}$ 和向量 $\boldsymbol{r}$。

(b) 找出最佳预测系数。

(c) 计算使用最佳系数得到的预测误差的方差。

8.29 推导出 $L = 8$ 时的劳埃德·马克斯判决-重建级别和均匀概率密度函数。

$$p(s) = \begin{cases} \dfrac{1}{2A}, & -A \leqslant s \leqslant A \\ 0, & \text{其他} \end{cases}$$

8.30 来自一所著名研究型医院的一位放射学家最近参加了一次医学会议，会上展示了一种可以通过标准 T1(1.544 Mbps)电话线路传输 4096×4096 大小的 12 比特数字化 X 光图像的系统。该系统使用一种渐进技术来传输压缩形式的图像，这种技术可在观察站首先重建出 X 射线图像的较好近似图像，然后逐步细化产生一幅无误差的显示。产生第一幅近似图像的数据传输要 5~6 秒的时间。在接下来的 1 分钟时间内，每隔 5~6 秒(平均)会进行一次细化，在重建的 X 射线图像中，第一次和最后一次细化有最大和最小的影响。这位物理学家对这套系统的印象很好，因为她可以使用第一幅 X 射线图像的近似开始进行诊断，而在生成无误差 X 射线重建图像时就可以完成诊断。在回办公室的途中，她拟定了一份报送给医院领导的采购意向。遗憾的是，医院正值预算紧张时期，因为最近雇用了一位年轻有为的电气工程专业毕业生。为了满足放射学家的要求，院方向年轻的工程师布置了设计这样一个系统的任务(院方认为设计和构造一个类似的自用系统成本会更低。医院现在拥有这种系统的某些部件，但传输原始 X 射线数据的时间要超过 2 分钟)。院方要求工程师在下午的工作会议上拿出一个初始框图。由于时间紧且手头只有一本《数字图像处理》教材(当然需要这本书)，工程师只能在概念上设计一个系统，以满足传输和相关的压缩要求。请构建该系统的概念框图，并详细说明你将建议的压缩技术。

8.31 证明由式(8.2-62)定义的基于提升的小波变换，与使用表 8.15 中系数实现的传统 FWT 滤波器组是等价的。依据 $\alpha, \beta, \gamma, \delta$ 和 K 定义滤波器系数。

8.32 计算一幅 JPEG-2000 编码图像的子带的量化步长。该图像采用导出式量化，对 $2LL$ 子带的尾数和指数分配 8 比特数据。

8.33 你如何在频率域将水印加到一幅图像上？

★8.34 以离散傅里叶变换为基础，设计一个不可见水印处理系统。

8.35 以离散小波变换为基础，设计一个不可见水印处理系统。

647 ~ 648

第9章　形态学图像处理

In form and feature, face and limb, I grew so like my brother.
That folks got taking me for him and each for one another.

Henry Sambrooke Leigh, Carols of Cockayne, The Twins

引言

形态学(morphology)一词通常表示生物学的一个分支，该分支主要研究动植物的形态和结构。这里，我们使用同一词语表示数学形态学的内容，将数学形态学作为工具从图像中提取表达和描绘区域形状的有用图像分量，如边界、骨架和凸壳等。我们对预处理或后处理的形态学技术也感兴趣，比如形态学过滤、细化和修剪等。

在下面几节中，我们将建立并说明数学形态学中的几个重要概念。这里介绍的许多概念可在 n 维欧氏空间 E^n 中用公式表达，然而，我们的兴趣一开始是二值图像，这种图像的各个分量是 Z^2 的元素(见2.4.2节)。在9.6节，我们的讨论将扩展到灰度图像。

本章的内容开始从输入和输出都是图像的图像处理，转变为输入是图像而输出是从这些图像中提取的属性的处理，详见1.1节的定义。形态学及与其相关数学概念，用于从图像中提取“内涵”。本书的余下章节中将探讨和应用其他方法。

9.1　预备知识

数学形态学的语言是集合论。同样，形态学为大量的图像处理问题提供一种一致且有力的方法。数学形态学中的集合表示图像中的对象。例如，在二值图像中，所有白色像素的集合是该图像的一个完整的形态学描述。在二值图像中，问题集合是二维整数空间 Z^2 的元素(见2.4.2节)，在该空间中，集合的每个元素都是一个多元组(二维向量)，这些多元组的坐标是图像中一个白色(或黑色，取决于事先的约定)像素的坐标(x, y)。前一章讨论过的灰度数字图像可以表示为其分量在空间 Z^3 中的集合。在这种情况下，集合中每个元素的两个分量提供一个像素的坐标，第三个分量则对应于其离散灰度值。更高维度空间中的集合可以包含其他的图像属性，如颜色和随时间变化的分量。

在继续阅读之前，请读者回顾一下2.4.2节和2.6.4节的内容。

除2.6.4节中的基本集合定义外，集合的反射和平移概念在形态学中用得也很广泛。一个集合 B 的反射表示为 $\hat{B}$，定义如下：

集合反射操作类似于空间卷积中执行的翻转(旋转)操作(见3.4.2节)。

649 ~ 650

$$\hat{B} = \{w \mid w = -b,\ b \in B\} \tag{9.1-1}$$

如果 B 是描述图像中物体的像素的集合(二维点)，则 $\hat{B}$ 是 B 中(x, y)坐标被$(-x, -y)$替代的点的集合。图 9.1(a)和(b)显示了一个简单的集合及其反射[1]。

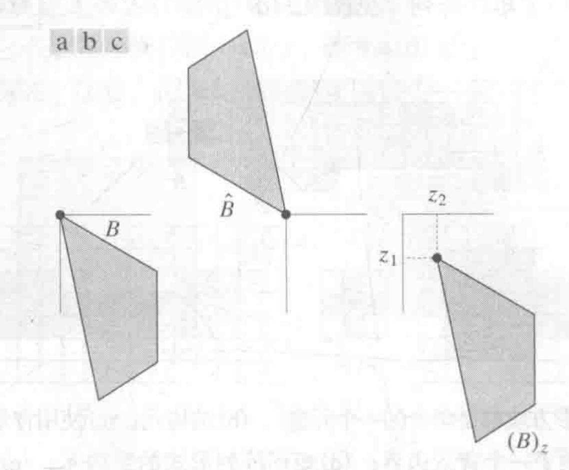


图 9.1　(a)一个集合；(b)该集合的反射；(c)距离为 z 的平移

集合 B 按照点 $z = (z_1, z_2)$ 表示为 $(B)_z$ 的平移定义如下：

$$(B)_z = \{c \mid c = b + z,\ b \in B\} \tag{9.1-2}$$

若 B 是描述图像中物体的像素集合，则 $(B)_z$ 是 B 中(x, y)坐标被 $(x + z_1, y + z_2)$ 替代的点的集合。图 9.1(c)使用来自图 9.1(a)的集合 B 说明了这一概念。

形态学中集合的反射和平移广泛用来表达基于结构元(SE)的操作：研究一幅图像中感兴趣特性所用的小集合或子图像。图 9.2 的第一行显示了结构元的几个例子，其中每个涂阴影的方块表示 SE 的一个成员。给定结构元中的一个位置是否是该 SE 集合的成员没有关系时，该位置用“×”来标记，表示一个“不关心”条件，如稍后在 9.5.4 节中定义的那样。除了元素是 SE 的成员的定义之外，还必须指定结构元的原点。图 9.2 中各种 SE 的原点由一个黑点指出(尽管将 SE 的中心放在其重心处很普遍，但通常原点的选择取决于具体问题)。当 SE 对称且未显示原点时，则假定原点位于对称中心处。

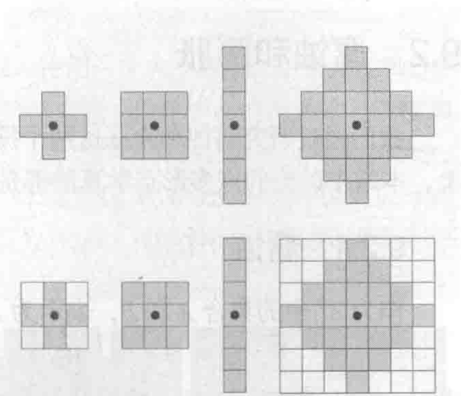


图9.2　第一行：结构元的例子；第二行：转换为矩形阵列的结构元。点表示结构元的中心

对图像操作时，我们要求结构元是矩形阵列。这是通过添加最小可能数量的背景元素(图 9.2 中所示的非阴影部分)形成一个矩形阵列来实现的。图 9.2 第二行中的第一个和最后一个 SE 说明了该过程。该行中的其他 SE 已是矩形形式。

为介绍在形态学中如何使用结构元，我们考虑图 9.3。图9.3(a)和(b)显示了一个简单的集合和一个结构元。如前一段提到的那样，计算机实现要求用添加背景元素的方法把 A 也转换为一个矩形阵列。当结构元的原点位于原始集合的边界上时，背景边界要大到足以容纳整个结构元(这类似于 3.4.2 节

---

① 对图像操作时，如图9.1中的集合，我们使用阴影来表示正考虑集合的成员的点(像素)。在处理二值图像时，感兴趣的集合是对应于物体的像素。我们使用白色像素来显示这些集合，而其他像素则显示为黑色。术语前景和背景通常分别用于表示图像中为物体和非物体定义的像素集合。

650 ~ 651

讨论的空间相关和卷积的填充操作)。在这种情况下，结构元的大小为3×3，原点位于中心，所以包围整个集合的一个元素的边界是足够的，如图9.3(c)所示。就像在图9.2中那样，必须使用最小可能数量的背景元素填充结构元，使它成为一个矩形阵列［见图9.3(d)］。

在后面的说明中，我们会添加足够的背景点来形成矩形阵列，但在含义明确的情形下，为简化图形，我们假定已进行了填充操作。

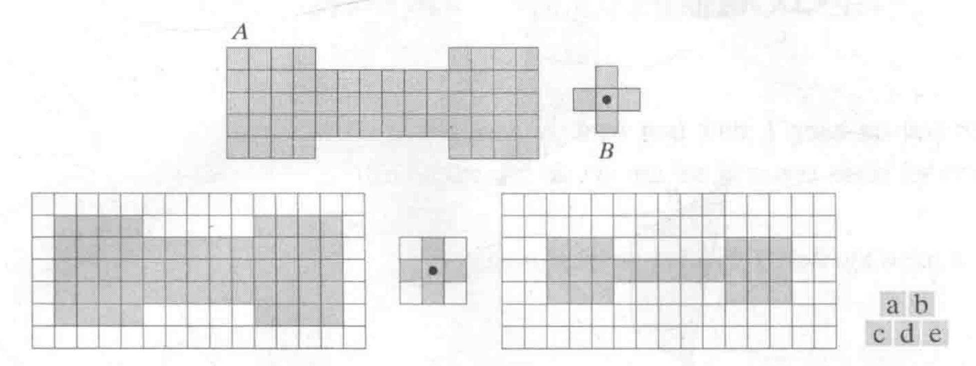


图9.3 (a)一个集合(每个阴影方块都是集合的一个元素)；(b)结构元；(c)使用背景元素填充集合形成的一个矩形阵列，并且提供了一个背景边界；(d)矩形阵列形式的结构元；(e)由结构元处理过的集合

假定我们定义一个用结构元B在集合A上的操作如下：通过让B在A上运行，以便B的原点访问A的每个元素，来创建一个新集合。在B的每个原点位置，如果B完全被A包含，则将该位置标记为新集合的一个成员(阴影所示)；否则，将该位置标记为非新集合的成员(非阴影所示)。图9.3(e)显示了这一操作的结果。我们看到，当B的原点位于A的边界元素上时，B的一部分将不再包含在A中，从而排除了B处在中心位置的点作为新集合的成员的可能。最终结果是集合的边界被腐蚀，如图9.3(e)所示。当我们使用“结构元包含在集合中”这样的术语时，我们明确地指出A和B的元素完全重叠。换句话说，虽然我们用包含阴影和非阴影的元素阵列来说明A和B，但在确定B是否包含于A中时，我们只考虑两个集合中的阴影元素。这些概念是下一节内容的基础，因此在继续学习之前全面理解图9.3中的概念很重要。

9.2 腐蚀和膨胀

我们通过研究腐蚀和膨胀这两个操作来开始形态学的讨论。这些操作是形态学处理的基础。事实上，本章中讨论的许多形态学算法都是以这两种原始操作为基础的。

9.2.1 腐蚀

作为Z^2中的集合A和B，表示为$A\ominus B$的B对A的腐蚀义为

$$A\ominus B=\{z\mid(B)_z\subseteq A\} \qquad (9.2\text{-}1)$$

表面上，该式指出B对A的腐蚀是一个用z平移的B包含在A中的所有的点z的集合。在下面的讨论中，假定集合B是一个结构元。式(9.2-1)是上节末尾讨论的图9.3(e)中的例子的数学公式。因为B必须包含在A中这一陈述等价于B不与背景共享任何公共元素，故可以将腐蚀表达为如下的等价形式：

$$A\ominus B=\{z\mid(B)_z\cap A^c=\varnothing\} \qquad (9.2\text{-}2)$$

652 ~ 653

式中，如2.6.4节所定义的那样，A^c是A的补集，$\varnothing$是空集。

图 9.4 显示了腐蚀的一个例子。A 和 B 的元素显示为阴影，背景显示为白色。图 9.4(c) 中的实线边界是 B 的原点进一步移动的界限，若超出该界限，会导致结构元不再完全包含于 A 中。这样，该边界内的点的轨迹(B 的原点位置)就构成了 B 对 A 的腐蚀。我们在图 9.4(c) 中使用阴影显示了该腐蚀的结果。记住，腐蚀是简单地满足式(9.2-1)或式(9.2-2)的 z 值的一个集合。图 9.4(c) 和(e) 中虚线所示的集合 A 的边界仅供参考，它不是腐蚀操作的一部分。图 9.4(d) 显示了一个拉长的结构元，图 9.4(e) 显示了该结构元对集合 A 的腐蚀。注意，原来的集合已被腐蚀为一条直线。

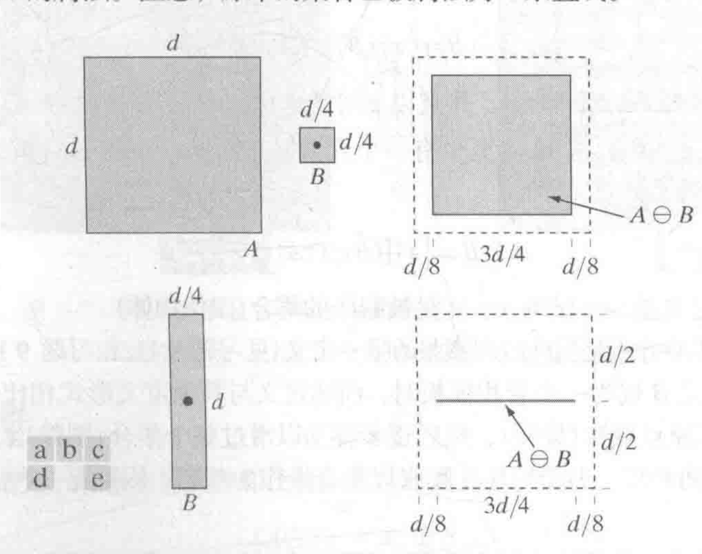


图 9.4　(a) 集合 A；(b) 方形结构元 B；(c) B 对 A 的腐蚀，如阴影部分所示；(d) 拉长的结构元；(e) 用拉长后的结构元 B 对 A 的腐蚀。(c) 和(e) 中的虚线边界是集合 A 的边界，仅供参考而显示

式(9.2-1)和式(9.2-2)不是腐蚀的唯一定义形式(见习题 9.9 和习题 9.10 中另外两种等价的定义形式)。然而，这些公式较之其他公式具有独特的优点，即把结构元 B 视为一个空间模板时(见 3.4.1 节)，它们更直观。

例 9.1　使用腐蚀去除图像的某些部分。

假设我们希望去掉图 9.5(a) 中连接中心区域到边界焊接点的线。使用一个大小为 11×11 且元素都是 1 的方形结构元腐蚀该图像，如图 9.5(b) 所示，大多数为 1 的线条都被去除。位于中心的两条垂直线被细化，但未被完全去除，原因是它们的宽度大于 11 个像素。把 SE 的大小改为 15×15，并再次腐蚀原图像，如图 9.5(c) 所示，所有的连线都去除了［另一种方法是，使用 11×11 的 SE 对图 9.5(b) 的图像再进行腐蚀］。增大结构元的尺寸甚至会消除更大的部件。例如，使用大小为 45×45 的结构元，可去除边界的焊接点，如图 9.5(d) 所示。

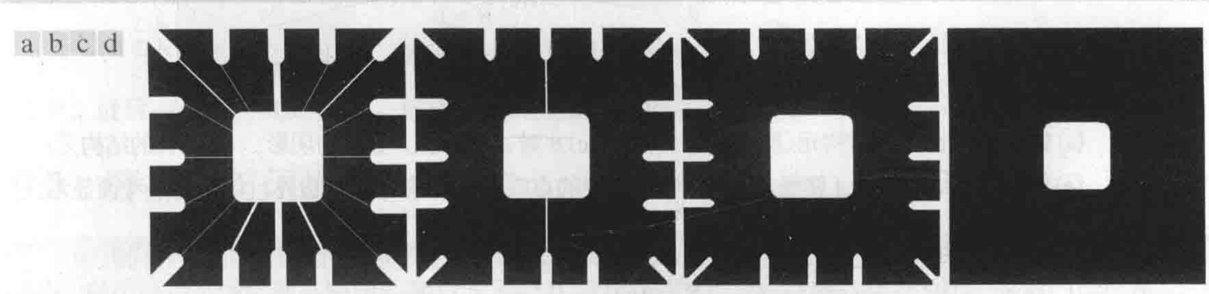


图 9.5　使用腐蚀去除图像中的部件：(a) 一幅大小为 486×486 的连线模板二值图像；(b)～(d) 分别使用大小为 11×11, 15×15 和 45×45 的结构元腐蚀的图像。SE 的元素都是 1

653～654

我们从这个例子看到，腐蚀缩小或细化了二值图像中的物体。事实上，我们可以将腐蚀视为形态学滤波操作，这种操作将小于结构元的图像细节从图像中滤除(去除)。在图 9.5 中，腐蚀执行了一个“线滤波”的功能。我们在 9.3 节和 9.6.3 节将会返回到形态学滤波这一概念。

9.2.2 膨胀

A 和 B 是 Z^2 中的集合，表示为 $A\oplus B$ 的 B 对 A 的膨胀定义为

$$A\oplus B=\left\{z\mid(\hat{B})_z\cap A\neq\varnothing\right\} \tag{9.2-3}$$

这个公式是以 B 关于它的原点的映像，并且以 z 对映像进行平移为基础(见图 9.1)。B 对 A 的膨胀是所有位移 z 的集合，这样，$\hat{B}$ 和 A 至少有一个元素是重叠的。根据这种解释，式(9.2-3)可以等价地写为

$$A\oplus B=\left\{z\mid[(\hat{B})_z\cap A]\subseteq A\right\} \tag{9.2-4}$$

与以前一样，我们假定 B 是一个结构元，A 是被膨胀的集合(图像物体)。

式(9.2-3)和式(9.2-4)并不是当前所用膨胀的唯一定义(见习题 9.11 和习题 9.12 中两种不同但等价的定义)。但在把结构元 B 视为一个卷积模板时，前述定义与其他定义形式相比优点明显，这种定义形式更直观。B 关于其原点翻转(旋转)，然后逐步移动以滑过整个集合(图像)A，这一基本过程类似于 3.4.2 节中介绍的空间卷积。但要记住，膨胀以集合操作为基础，因此是一种非线性操作，而卷积是一种线性操作。

与腐蚀不同，腐蚀是一种收缩或细化操作，膨胀则会“增长”或“粗化”二值图像中的物体。这种特殊的方式和粗化的宽度由所用结构元来控制。图 9.6(a)显示了与图 9.4 中所用集合相同的集合，图 9.6(b)显示了一个结构元(在这种情况下，$\hat{B}=B$，因为 SE 关于其原点对称)。作为参考，图 9.6(c)中的虚线显示了原始集合，实线显示了一个界限，$\hat{B}$ 的原点进一步移动 z，若超出这一界限，则导致 $\hat{B}$ 和 A 的交集为空。因此，处在该边界上或该边界内的所有点就构成了 B 对 A 的膨胀。图 9.6(d)显示了一个设计用来实现垂直方向膨胀比水平方向膨胀更多的结构元，图 9.6(e)则显示了使用这个结构元进行膨胀的结果。

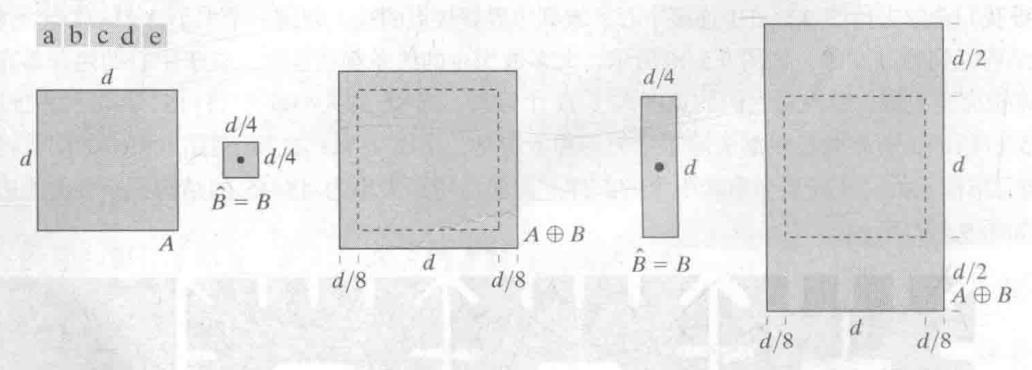


图 9.6 (a)集合 A；(b)方形结构元(黑点表示原点)；(c)B 对 A 的膨胀，显示为阴影；(d)拉长的结构元；(e)使用这个结构元对 A 膨胀。图(c)和图(e)中的点线的边是集合 A 的边界，仅作为参考而显示

例 9.2 膨胀的说明。

最简单的膨胀应用之一是连接裂缝。图 9.7(a)显示了与图 4.49 相同的带有断裂的字符的图像，当时在图 4.49 中是用低通滤波进行连接的。已知断裂的最大长度为两个像素。图 9.7(b)显示了能够修复这些断裂的结构元(注意，替代阴影，我们用 1 表示 SE 的元素，而用 0 表示背景；这是因为现在 SE 被当作

655 ~ 656

一幅子图像来处理，而不是一幅图形)。图9.7(c)显示了使用这个结构元对原图像进行膨胀后的结果。裂缝已被连接。形态学方法较之我们在图4.49中用于连接断裂的低通滤波方法的一个直接优点是，形态学方法可在一幅二值图像中直接得到结果。另一方面，低通滤波方法则从一幅二值图像开始，生成一幅灰度图像，它需要用一个阈值函数将灰度图像转换为二值图像。

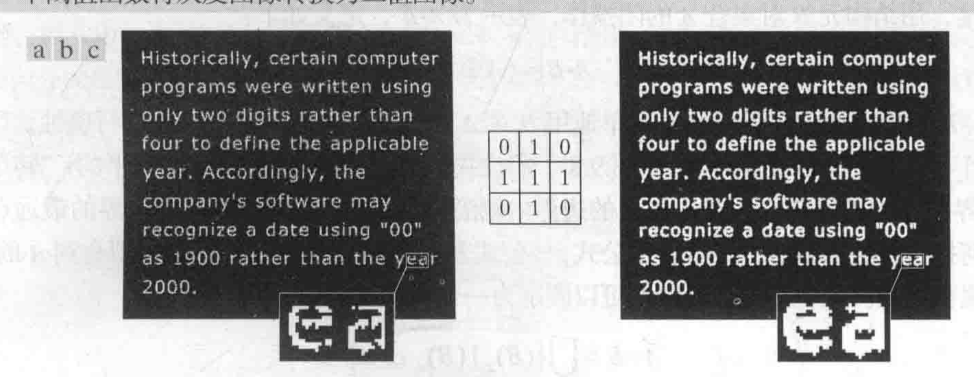


图 9.7　(a)具有断裂字符的低分辨率样品文本(见放大的视图)；(b)结构元；(c)图(b)对(a)的膨胀。断裂线段被连起来了

9.2.3　对偶性

膨胀和腐蚀彼此关于集合求补运算和反射运算是对偶的，即

$$(A\ominus B)^c = A^c \oplus \hat{B} \tag{9.2-5}$$

和

$$(A\oplus B)^c = A^c \ominus \hat{B} \tag{9.2-6}$$

式(9.2-5)指出，B 对 A 的腐蚀是 $\hat{B}$ 对 A^c 的膨胀的补集，反之亦然。当结构元关于其原点对称时(通常如此)，因为 $\hat{B}=B$，故对偶性特别有用。这样，我们可以用相同的结构元简单地使用 B 膨胀图像的背景(即膨胀 A^c)，对该结果求补即可得到 B 对该幅图像的腐蚀。类似的说明适用于式(9.2-6)。

为了说明建立形态学表达式有效性的一种典型方法，我们正式证明式(9.2-5)的有效性。由腐蚀的定义，我们有

$$(A\ominus B)^c = \left\{z \mid (B)_z \subseteq A\right\}^c$$

若集合 $(B)_z$ 包含在集合 A 中，则 $(B)_z \cap A^c = \varnothing$，在这种情况下，上式变为

$$(A\ominus B)^c = \left\{z \mid (B)_z \cap A^c = \varnothing\right\}^c$$

但是，满足 $(B)_z \cap A^c = \varnothing$ 的 z 的集合的补集是满足 $(B)_z \cap A^c \neq \varnothing$ 的 z 的集合，因此有

$$(A\ominus B)^c = \left\{z \mid (B)_z \cap A^c \neq \varnothing\right\} = A^c \oplus \hat{B}$$

其中，最后一步来自式(9.2-3)。证毕。采用类似的推理可以证明式(9.2-6)(见习题 9.13)。

9.3　开操作与闭操作

如我们所见的那样，膨胀会扩大一幅图像的组成部分，而腐蚀则会缩小一幅图像中的组成部分。本节讨论另外两个重要的形态学操作：开操作与闭操作。开操作一般会平滑物体的轮廓、断开较窄的狭颈并消除较细的突出物。闭操作同样也会平滑轮廓的一部分，但与开操作相反，它通常会弥合较窄的间断和细长的沟壑，消除较小的孔洞，填补轮廓线中的断裂。

656 ~ 657

结构元B对集合A的开操作，表示为$A\circ B$，定义如下：

$$A\circ B=(A\ominus B)\oplus B \tag{9.3-1}$$

因此，B对A的开操作就是B对A的腐蚀，紧接着用B对结果进行膨胀。

类似地，用结构元B对集合A的闭操作，表示为$A\bullet B$，定义如下：

$$A\bullet B=(A\oplus B)\ominus B \tag{9.3-2}$$

上式说明，B对集合A的闭操作就是简单地用B对A膨胀，紧接着用B对结果进行腐蚀。

开操作有一个简单的几何解释(见图 9.8)。假设我们把结构元B视为一个(扁平的)“转球”。然后，$A\circ B$的边界由B中的点建立：当B在A的边界内侧滚动时，B所能到达的A的边界的最远点。开操作的这种几何拟合特性导致了一个集合论公式，该公式表明B对A的开操作是通过拟合到A的B的所有平移的并集得到的。也就是说，开操作可以表示为一个拟合处理：

$$A\circ B=\bigcup\{(B)_z \mid (B)_z\subseteq A\} \tag{9.3-3}$$

其中$\cup\{\}$表示大括号中所有集合的并集。

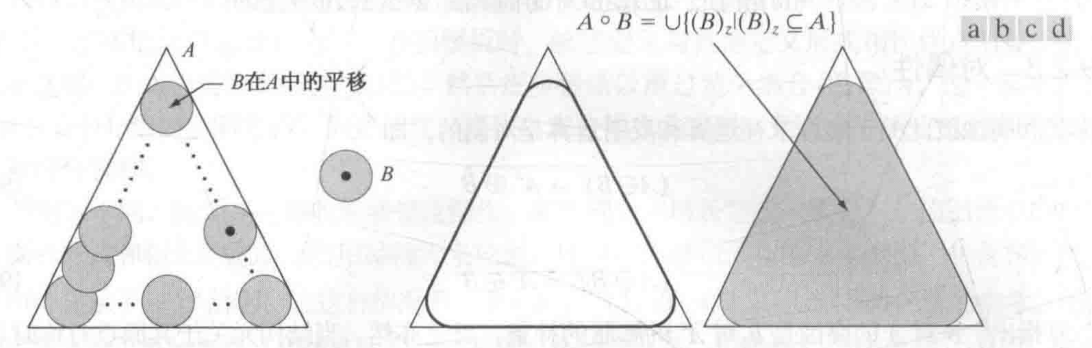


图 9.8 (a)结构元B沿集合A的内侧边界滚动(黑点表示B的原点)；(b)结构元；(c)粗线是开操作的外部边界；(d)完全的开操作(阴影部分)。为清楚起见，在图(a)中我们未对A加阴影

除了我们现在是在边界的外侧滚动B(见图9.9)之外，闭操作有类似的几何解释。如下面所讨论的那样，开操作和闭操作彼此对偶，所以闭操作在边界外侧滚动球体是意料之中的事情。从几何上讲，当且仅当对包含w的$(B)_z$进行的任何平移都有$(B)_z\cap A\neq\varnothing$时，点$w$才是$A\bullet B$的一个元素。图9.9说明了闭操作这一基本的几何性质。

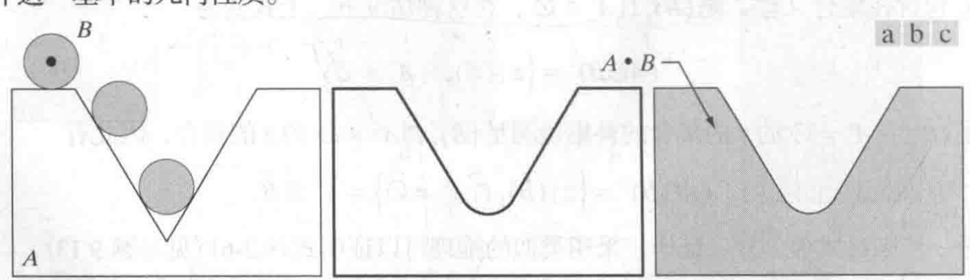


图 9.9 (a)结构元B沿集合A的外侧边界滚动；(b)粗线是闭操作的外部边界；(c)完全的闭操作(阴影部分)。为清楚起见，在图(a)中我们未对A加阴影

658 ~ 659

例 9.3 形态学开操作和闭操作的简单说明。

图 9.10 进一步说明了开操作和闭操作。图 9.10(a)显示了一个集合A，图 9.10(b)显示了腐蚀处理期间一个圆盘形结构元的各个位置。腐蚀完成后，得到图 9.10(c)所示的分离图形。注意，两个主要

部分之间的连接消失。连接部分的宽度与结构元的直径相比要细；也就是说，集合的这部分不能完全包含结构元，因此违反了式(9.2-1)的条件。该物体最右边的两个部分也是如此。圆盘无法拟合的突出部分已被消除。图 9.10(d) 显示了对腐蚀后的集合进行膨胀的处理，图 9.10(e) 显示了开操作的最终结果。注意，方向向外的角变圆，而方向向内的角则未受影响。

类似地，图 9.10(f) 到 (i) 显示了使用同一结构元对 A 进行闭操作的结果。我们注意到方向向内的角变圆，而方向向外的角则保持不变。在 A 的边界上，最左边的突入部分的尺寸明显减小，因为在这个位置上圆盘无法拟合。还要注意使用圆盘形结构元对集合 A 进行开操作和闭操作所得到的物体的各个部分都变得平滑了。

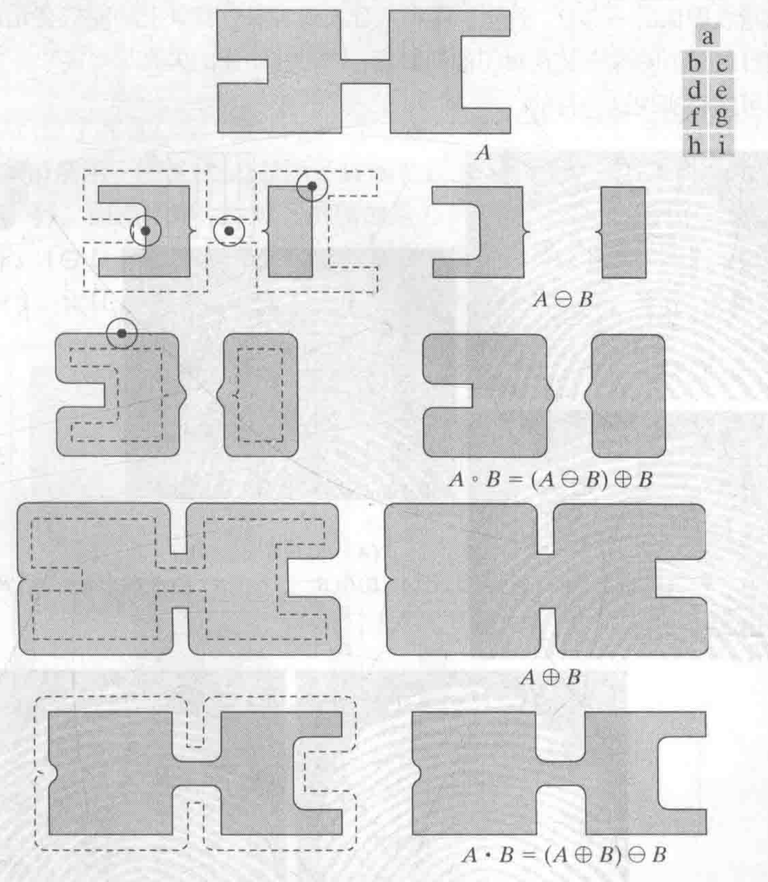


图 9.10　形态学开操作和闭操作。结构元是在图(b)中的各个位置显示的小圆。为清楚起见，SE 未加阴影。黑点是结构元的中心

如同膨胀和腐蚀的情形那样，开操作和闭操作彼此关于集合求补和反射也是对偶的，即

$$(A \bullet B)^c = (A^c \circ \hat{B}) \tag{9.3-4}$$

和

$$(A \circ B)^c = (A^c \bullet \hat{B}) \tag{9.3-5}$$

我们将该结果的证明作为练习留给读者(见习题 9.14)。

开操作满足下列性质：

(a) $A \circ B$ 是 A 的一个子集(子图像)。

(b) 如果 C 是 D 的一个子集，则 $C \circ B$ 是 $D \circ B$ 的一个子集。

(c) $(A \circ B) \circ B = A \circ B$。

659~660

类似地，闭操作满足下列性质：

(a) A 是 $A \cdot B$ 的一个子集(子图像)。

(b) 如果 C 是 D 的一个子集，则 $C \cdot B$ 是 $D \cdot B$ 的一个子集。

(c) $(A \cdot B) \cdot B = A \cdot B$ 。

注意，由两种情况下的条件(c)可知，算子应用一次后，一个集合的多次开操作或闭操作没有影响。

例 9.4 针对形态学滤波使用开操作和闭操作。

形态学操作可用于构建与第 3 章中讨论的空间滤波概念相类似的滤波器。图 9.11(a)中的二值图像显示了被噪声污染的指纹图像的一部分。这里，噪声本身表现为黑色背景上的随机亮元素和指纹较亮部分上的暗元素。我们的目的是消除噪声及其对印刷的影响，同时使图像的失真尽可能小。由开操作后紧跟闭操作的形态学滤波器可用来实现这一目的。

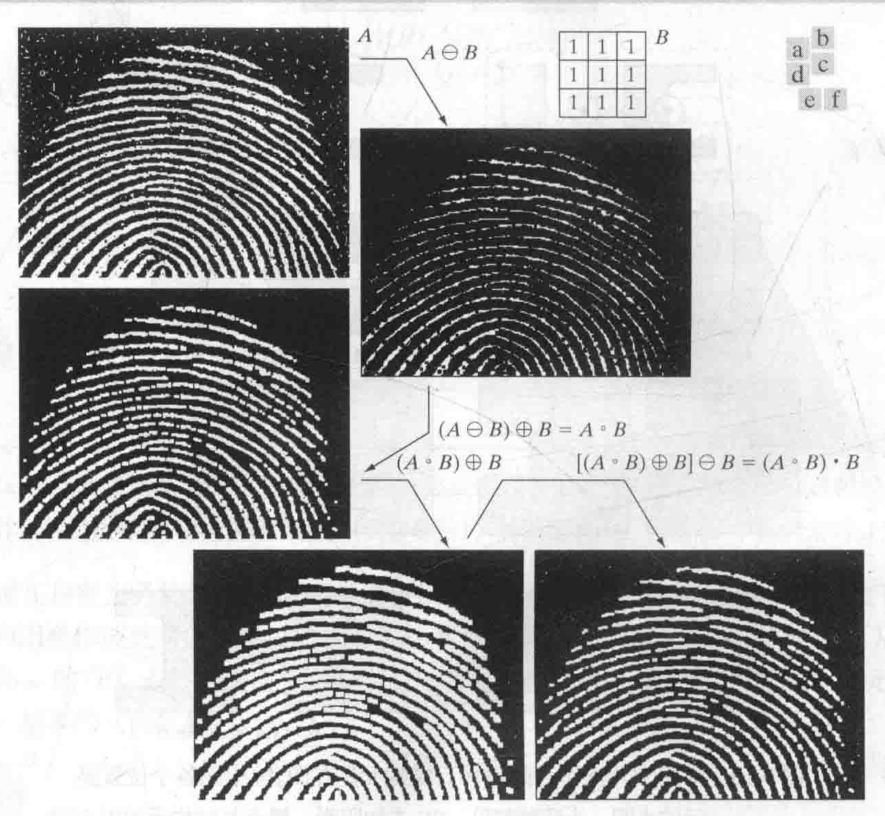


图 9.11 (a)噪声图像；(b)结构元；(c)腐蚀后的图像；(d)A 的开操作；(e)开操作的膨胀；(f)开操作的闭操作(原图由美国国家标准和技术研究所提供)

图 9.11(b)显示了所用的结构元。图 9.11 的余下部分按步骤顺序显示了滤波操作。图 9.11(c)显示了使用结构元对 A 进行腐蚀的结果。背景噪声在开操作的腐蚀阶段被完全消除，因为这种情况下的所有噪声分量都比结构元小。包含在指纹中的噪声元素(黑点)的尺寸实际上却增大了。原因是当物体被腐蚀时，这些元素在尺寸增大的内部边界。这种增大可通过对图 9.11(c)执行膨胀来抵消。图 9.11(d)显示了该结果，指纹中包含的噪声分量的尺寸被减小，或完全被消除。

刚才描述的两个操作构成了 B 对 A 的开操作。我们注意到，在图 9.11(d)中，开操作的实际效果实质上是从背景和指纹本身中消除所有噪声。然而，在指纹纹路间却产生了新断裂。为防止这种不希望的影响，我们在开操作上执行膨胀，如图 9.11(e)所示。大部分断裂得以恢复，但纹路却变粗了，这是可

660 ~ 661

由腐蚀来弥补的一种情形。图 9.11(f)所示的结果构成了图 9.11(d)的开操作的闭操作。最后的结果是噪声斑点清除得相当干净，但这种方法有缺点，即有些指纹纹路未被完全修复，并还有间断。对这种情况也并非毫无指望，只是因为在保持连续性方面未插入任何条件而已(我们将在例 9.8 中再次讨论这一问题，并在 11.1.7 节中论证处理这一问题的方法)。

9.4　击中或击不中变换

形态学击中或击不中变换是形状检测的基本工具。我们将借助于图 9.12 来说明这一概念。图 9.12 显示了一个由三种形状(子集)组成的集合 A，三种形状分别由 C, D 和 E 表示。图 9.12(a)到(c)中的阴影部分表示原始集合，而图 9.12(d)和(e)中的阴影部分显示形态学操作的结果。目的是找到三种形状之一的位置，譬如说 D 的位置。

令每种形状的原点位于其重心处。设 D 被一个小窗口 W 包围。关于 W 的 D 的局部背景定义为差集$(W-D)$，如图 9.12(b)所示。图 9.12(c)显示了 A 的补集，后面将会用到它。图 9.12(d)显示了用 D 对 A 的腐蚀(虚线仅供参考)。回顾可知，D 对 A 的腐蚀是 D 的原点位置的集合，这样，D 就完全包含在 A 中。换一种方法解释，$A\ominus D$ 在几何上可视为 D 的原点的所有位置的集合，在每个这样的位置，D 找到 A 中的一个匹配(击中)。记住，在图 9.12 中，A 仅由三个不相交的集合 C, D 和 E 组成。

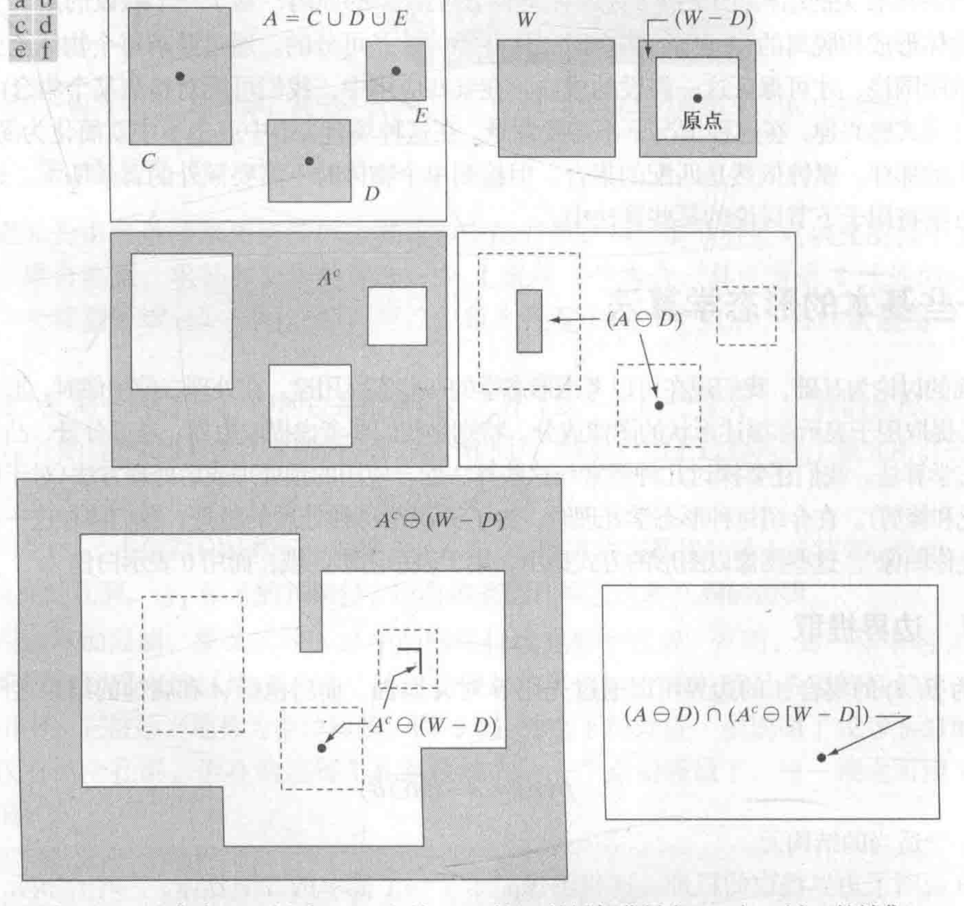


图 9.12　(a)集合 A；(b)窗口 W 和关于 W 的 D 的局部背景$(W-D)$；(c)A 的补集；(d)D 对 A 的腐蚀；(e)$(W-D)$对 A^c 的腐蚀；(f)(d)和(e)的交集，如所希望的那样，该交集显示了 D 的原点的位置。黑点是 C, D 和 E 的原点

661 ~ 663

图9.12(e)显示了局部背景集合$(W-D)$对A的补集的腐蚀。图9.12(e)中的外部阴影区域是腐蚀的部分。我们从图9.12(d)和(e)注意到,D精确地拟合A内部的位置集合是由D对A的腐蚀和$(W-D)$对A^c的腐蚀的交集，如图9.12(f)所示。这个交集正好是我们要寻找的位置。换句话说，如果B表示由D及其背景组成的集合，则B在A中的匹配（或匹配集合），表示为$A\circledast B$，是

$$A\circledast B=(A\ominus D)\cap[A^c\ominus(W-D)] \tag{9.4-1}$$

我们可以通过令$B=(B_1,B_2)$对这种表示法稍微做些推广，其中B_1是由与一个目标相联系的B的元素构成的集合，B_2是由与相应背景相联系的B的元素构成的集合。根据前面的讨论，$B_1=D, B_2=(W-D)$。用这种表示方法，式(9.4-1)变为

$$A\circledast B=(A\ominus B_1)\cap(A^c\ominus B_2) \tag{9.4-2}$$

因此，集合$A\circledast B$包含了所有的原点，同时，在这些原点处，B_1在A中找到了一个匹配(击中)，B_2在A^c中也找到了一个匹配。使用式(2.6-19)中给出的差集的定义，及式(9.2-5)中给出的腐蚀和膨胀间的对偶关系，可将式(9.4-2)写为

$$A\circledast B=(A\ominus B_1)\cap(A\oplus\hat{B}_2) \tag{9.4-3}$$

然而，式(9.4-2)更为直观。我们将上述三个公式称为*形态学击中或击不中变换*。

使用与物体有关的结构元B_1和与背景有关的结构元B_2的原因，基于一个假设的定义——仅当两个或多个物体形成相脱离的(断开的)集合时，这些物体才是可分的。通过要求每个物体至少被一个像素宽的背景所围绕，才可保证这一假设的成立。在某些应用中，我们可能对检测某个集合内由1和0组成的某些模式感兴趣，在这种情况下不需要背景。在这种场合，击中或击不中就简化为简单的腐蚀。如前面指出的那样，腐蚀依然是匹配的集合，但检测单个物体时不需要额外的背景匹配。这种简化的模式检测方案将用于下节讨论的某些算法中。

9.5 一些基本的形态学算法

以前面的讨论为基础，我们现在可以考虑形态学的一些实际用途。在处理二值图像时，形态学的主要应用之一是提取用于表示和描述形状的图像成分。特别是我们要考虑提取边界、连通分量、凸壳和区域的骨架的形态学算法。我们还要探讨几种经常与这些算法配合使用的预处理或后处理方法(对于区域填充、细化、粗化和修剪)。在介绍每种形态学处理时，为了弄清楚每种处理的机理，我们将在这一小节中广泛地使用“迷你图像”。这些图像以图形的方式显示，用1表示阴影区域，而用0表示白色。

9.5.1 边界提取

表示为$\beta(A)$的集合A的边界可以通过先用B对A腐蚀，而后执行A和腐蚀的结果之间的差集得到，即

$$\beta(A)=A-(A\ominus B) \tag{9.5-1}$$

式中B是一个适当的结构元。

图9.13说明了边界提取的机理。这幅图像显示了一个简单的二值物体、一个结构元B和使用式(9.5-1)得到的结果。尽管图9.13(b)中的结构元最常用，但它绝对不是唯一的。例如，使用由1组成的大小为5×5的结构元，将得到2~3个像素宽的边界。

> 从现在开始，我们将不再显式地给出边界填充。

663~664

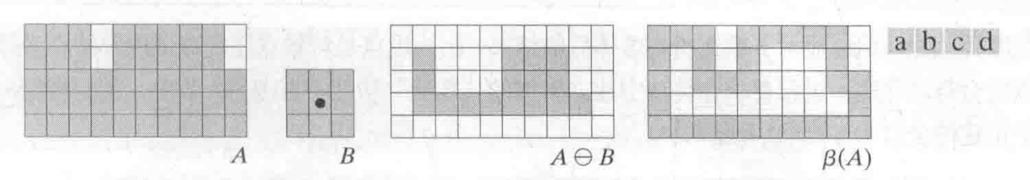


图 9.13　(a) 集合 A；(b) 结构元 B；(c) B 对 A 的腐蚀；(d) 由 A 和其腐蚀间的差集给出的边界

例 9.5　用形态学处理提取边界。

图 9.14 进一步说明了式 (9.5-1) 和由 1 组成的 3×3 结构元的用途。如本章中的所有二值图像那样，二进制值 1 显示为白色，二进制值 0 显示为黑色，因此该结构元的元素 1 也被当作白色来处理。由于所用结构元的尺寸，图 9.14(b) 中的边界宽度为 1 个像素。

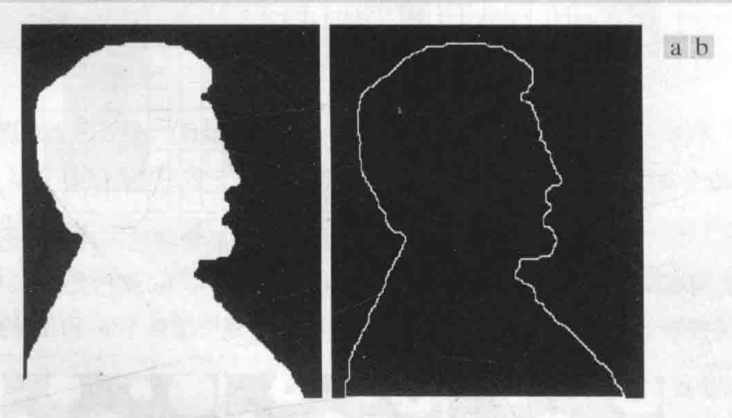


图 9.14　(a) 一幅简单的二值图像，其中 1 表示白色；(b) 使用式 (9.5-1) 和图 9.13(b) 中的结构元得到的结果

9.5.2　孔洞填充

孔洞定义为由前景像素相连接的边界所包围的背景区域。本节将针对填充图像中的孔洞，开发一种基于集合膨胀、求补和交集的算法。令 A 表示一个集合，其元素是 8 连通的边界，每个边界包围一个背景区域（即一个孔洞）。当给定每个孔洞中的一个点后，目的就是用 1 填充所有的孔洞。

除了每个孔洞中对应于 X_0 中的位置给定的点之外（这一点已经置为 1），我们从形成一个由 0 组成的阵列 X_0 开始（该阵列与包含 A 的阵列的大小相同）。然后，如下过程将用 1 填充所有孔洞：

$$X_k = (X_{k-1} \oplus B) \cap A^c \quad k = 1, 2, 3, \cdots \tag{9.5-2}$$

式中 B 是图 9.15(c) 中的对称结构元。如果 $X_k = X_{k-1}$，则算法在迭代的第 k 步结束。然后，集合 X_k 包含所有被填充的孔洞。X_k 和 A 的并集包含所有填充的孔洞及这些孔洞的边界。

如果左边不加限制，那么式 (9.5-2) 中的膨胀将填充整个区域。然而，每一步中与 A^c 的交集操作将把结果限制到感兴趣区域内。这是我们如何制约形态学处理以满足希望特性的第一个例子。在当前应用中，它被适当地称为条件膨胀。图 9.15 的余下部分进一步说明了式 (9.5-2) 的机理。尽管该例中仅有一个孔洞，但在给定每个孔洞区域内的一个点的假设下，这一概念可用于任何有限数量的孔洞。

例 9.6　形态学孔洞填充。

图9.16(a) 显示了一幅由内部带有黑色点的白色圆圈组成的图像。这样的图像可以通过将包含磨光的球体（如滚珠）的场景用阈值处理分为两个层次而得到。球体内部的黑点可能是反射的结果。我们的目的是通过孔洞填充来消除这些反射。图 9.16(a) 显示了在一个球体中选择的一个点，图 9.16(b) 显示了填充一部分

664 ~ 666

的结果。最后，图9.16(c)显示了填充所有球体后的结果。因为黑点是背景点还是球体内部的点必须是已知的，所以完全自动化这一过程要求在算法中建立附加的“智能”功能。在 9.5.9 节中，我们将给出一种基于形态学重建的全自动方法(见习题 9.23)。

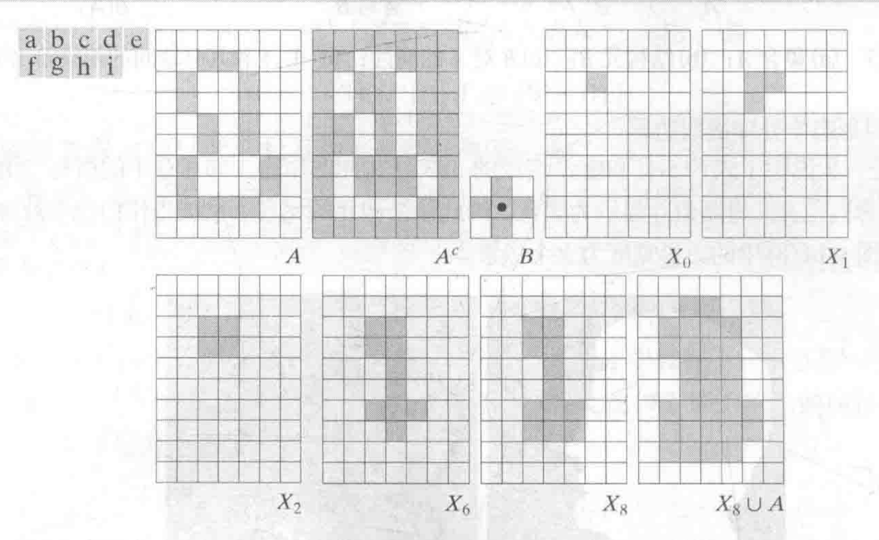


图 9.15 孔洞填充：(a)集合 A(显示为阴影)；(b)A 的补集；(c)结构元 B；(d)边界内的初始点；(e)~(h)式(9.5-2)的各个步骤；(i)最终结果［(a)和(h)的并集］

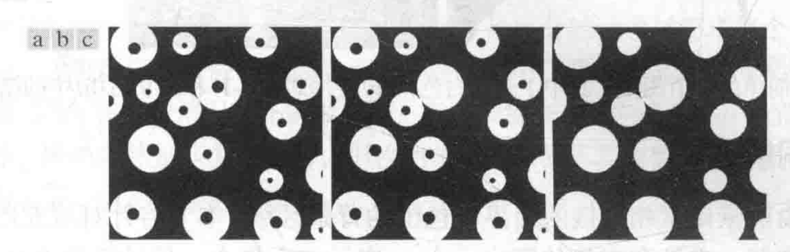


图 9.16 (a)二值图像(区域内部的白点是孔洞填充算法的起始点)；(b)填充该区域后的结果；(c)填充所有孔洞后的结果

9.5.3 连通分量的提取

连通性和连通分量的概念已在 2.5.2 节中介绍过。从二值图像中提取连通分量是许多自动图像分析应用的核心。令 A 是包含一个或多个连通分量的集合，并形成一个阵列 X_0 (该阵列的大小与包含 A 的阵列的大小相同)，除了在对应于 A 中每个连通分量的一个点的每个已知位置处我们已置为 1(前景值)外，该阵列的所有其他元素均为 0(背景值)。如下迭代过程可完成这一目的：

$$X_k = (X_{k-1} \oplus B) \cap A \quad k = 1, 2, 3, \cdots \tag{9.5-3}$$

式中，B 是一个适当的结构元(如图9.17所示)。当 $X_k = X_{k-1}$ 时，迭代过程结束，X_k 包含输入图像中的所有连通分量。注意式(9.5-3)与式(9.5-2)的相似性，唯一的差别是用 A 代替了 A^c。这并不奇怪，因为此处我们正在寻找前景点，而在 9.5.2 节的目的是寻找背景点。

图9.17说明了式(9.5-3)的机理，$k = 6$ 时即可收敛。注意，所用结构元的形状在像素间是基于 8 连通的。如果我们用图 9.15 中的 SE，它是基于 4 连通的，朝向图像底部的连通分量的最左侧元素将不会被检测到，因为对图的其余部分它是 8 连通的。如孔洞填充算法那样，假定在每个连通分量内都已知一个点，式(9.5-3)对于任何在 A 中的有限数量的连通分量都是可用的。

关于事先不要求知道每个连通分量中的一个点的算法，请参阅习题 9.24。

667
~
668

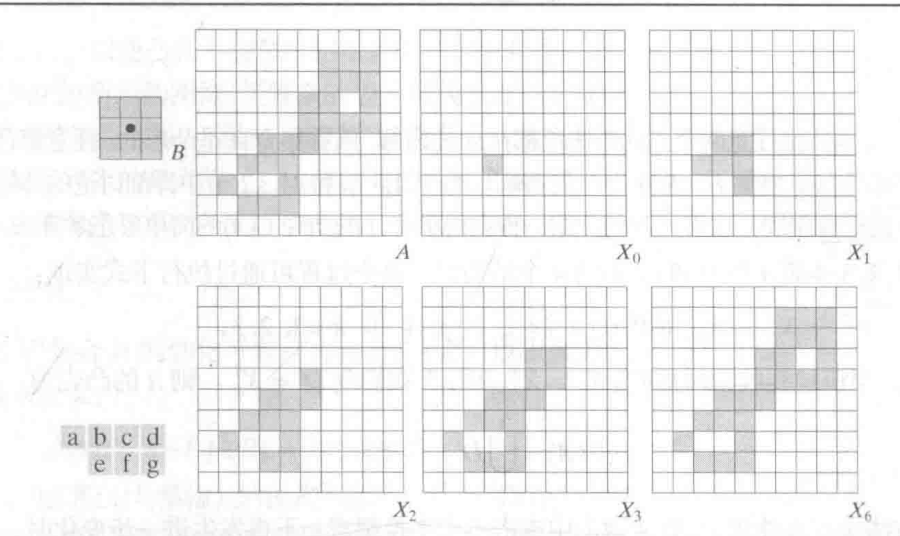


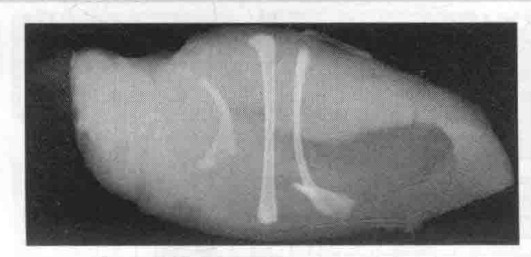
图 9.17　提取连通分量：(a)结构元；(b)包含有一个连通分量的集合的阵列；(c)在连通分量的区域中包含一个 1 的初始阵列；(d)~(g)式(9.5-3)的各个迭代步骤

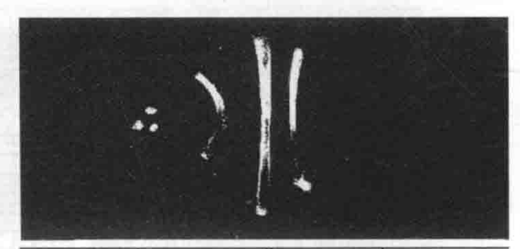
例 9.7　使用连通分量检测包装食品中的外来物。

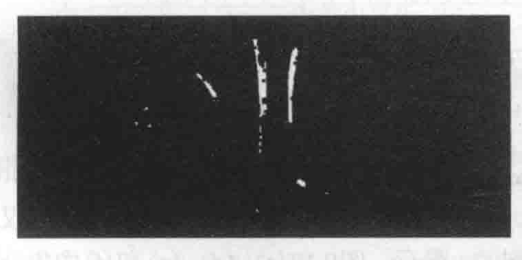
连通分量经常用于自动检测。图 9.18(a)显示了一幅含有碎骨的鸡胸 X 射线图像。经过处理的食品在包装和/或运送之前能检测出这样的物体是很有意义的。在这种特殊情况下，骨骼的密度使得它们的正常灰度值与背景不同。这就使得使用单一阈值(阈值处理在 3.1 节介绍过，10.3 节中将进行更详细的讨论)将骨骼从背景中提取出来非常简单。结果是图 9.18(b)所示的二值图像。

这幅图中的最显著特征是保留下来的点聚集为物体(骨头)，而不是彼此孤立的毫无关系的点。我们可以确定，只有具有“有效”尺寸的物体才能对经阈值处理的图像进行腐蚀后而保留下来。在这个例子中，我们将“有效”尺寸的物体定义为：使用元素为 1、大小为 5×5 的结构元腐蚀图像后保留下来的任何物体。图 9.18(c)中显示了腐蚀的结果。下一步是分析保留下来的物体的尺寸。我们通过提取图像中的连通分量来标记(识别)这些物体。图 9.18(d)中的表格列出了提取的结果。总共有 15 个连通分量，其中 4 个尺寸较大。这足以确定包含在原图像中的重要但不希望的物体。如果需要，可使用第11章讨论的技术进一步描述其特性(比如物体的形状)。

a b
c d







连通分量	连通分量中的像素数	连通分量	连通分量中的像素数
01	11	09	7
02	9	10	11
03	9	11	11
04	39	12	9
05	133	13	9
06	1	14	674
07	1	15	85
08	743		

图 9.18　(a)含有碎骨的鸡肉 X 射线图像；(b)经阈值处理后的图像；(c)使用元素为 1、大小为 5×5 的结构元腐蚀后的图像；(d)图(c)的连通分量中的像素数(图像由 NTB Elektronische Geraete GmbH, Diepholz, Germany, www.ntbxray.com 提供)

668
~
669

9.5.4 凸壳

如果在集合 A 内连接任意两个点的直线段都在 A 的内部，则称集合 A 是凸形的。任意集合 S 的凸壳 H 是包含于 S 的最小凸集。差集 $H-S$ 称为 S 的凸缺。如 11.1.6 节和 11.3.2 节中详细讨论的那样，凸壳和凸缺对于物体描绘是很有用的。这里，我们介绍一种获得集合 A 的凸壳 $C(A)$的简单形态学算法。

令 B^i , $i=1,2,3,4$ 表示图 9.19(a)中的 4 个结构元。这个过程可通过执行下式实现：

$$X_k^i=(X_{k-1}\circledast B^i)\cup A \quad i=1,\ 2,\ 3,\ 4 \ \text{和}\ k=1,2,3,\cdots \tag{9.5-4}$$

式中，$X_0^i=A$。当该过程收敛时(即当 $X_k^i=X_{k-1}^i$ 时)，我们令 $D^i=X_k^i$。则 A 的凸壳为

$$C(A)=\bigcup_{i=1}^{4}D^i \tag{9.5-5}$$

换句话说，该方法由反复使用 B^1 对 A 做击中或击不中变换组成；不再发生进一步变化时，我们执行与 A 的并集运算，结果称为 D^1。这一过程使用 B^2 重复(应用于 A)，直到不发生进一步的变化，如此往复。得到的 4 个 D 的并集组成了 A 的凸壳。注意，我们使用不需要背景匹配的击中或击不中变换的简化实现，如 9.4 节末尾所讨论的那样。

图 9.19 说明了式(9.5-4)和式(9.5-5)给出的过程。图 9.19(a)显示了用于提取凸壳的结构元。每个结构元的原点均位于其中心处。“×”项表示“不考虑”的条件。这意味着，如果结构元模板下 A 的 3×3 区域在该位置匹配模板的模式，则说结构元在 A 中找到了一个匹配。对于一个特殊的模板，当 A 中的这个 3×3 区域的中心为 0 时，阴影模板元素下的 3 个像素为 1 时，才会出现模式匹配。不必顾及这个 3×3 区域内其他像素的值。此外，图 9.19(a)中的符号 B^i 是由 B^{i-1} 顺时针旋转 90°得到的。

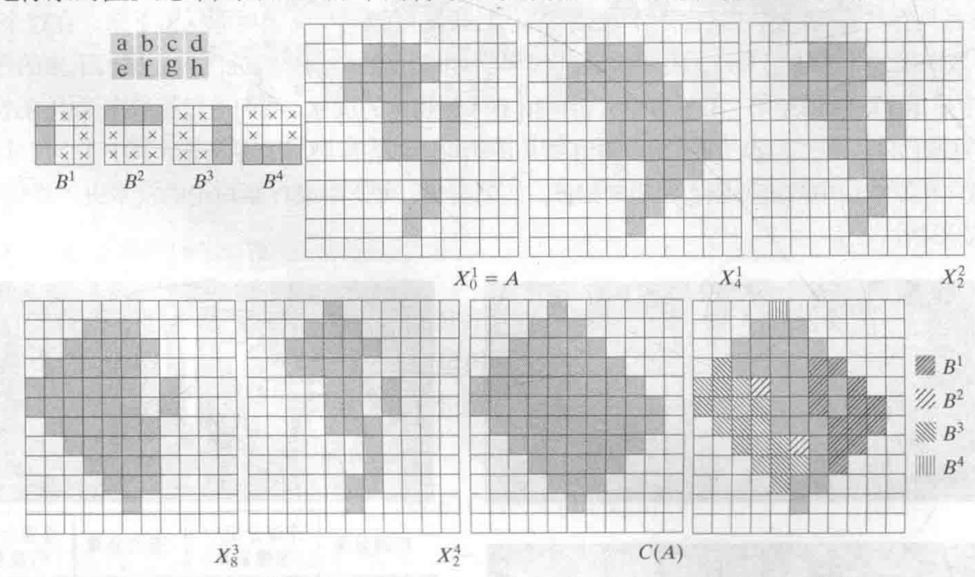


图 9.19 (a)结构元；(b)集合 A；(c)~(f)使用图(a)中的结构元得到的收敛结果；(g)凸壳；(h)显示了每个结构元的贡献的凸壳

图 9.19(b)显示了一个寻找凸壳的集合 A。从 $X_0^1=A$ 开始，式(9.5-4)经过 4 次迭代后，得到图 9.19(c)中的集合。然后令 $X_0^2=A$，并再次使用式(9.5-4)，得到图9.19(d)中的集合(在这种情况下，仅经过两步即可收敛)。接下来的两个结果是以同样的方式得到的。最后，图9.19(c)，(d)，(e)和(f)中集合形成的并集得到图 9.19(g)显示的凸壳。每个结构元的贡献突出显示在图 9.19(h)所示的组合集合中。

669 ~ 670

上述过程的一个明显的缺点是，凸壳可能超出确保凸性所需的最小尺寸。减少这种影响的一种简

单方法是限制生长，以便凸壳不会超过初始点集在水平和垂直方向上的尺寸。对图9.19中的例子施加这种限制产生图9.20所示的图像。更复杂的边界可用于进一步限制具有更多细节的图像的生长。例如，我们可以沿垂直、水平和对角线方向使用原始点集的最大尺寸。这样的改进所付出的代价是，额外增加了算法的复杂性和计算需求。

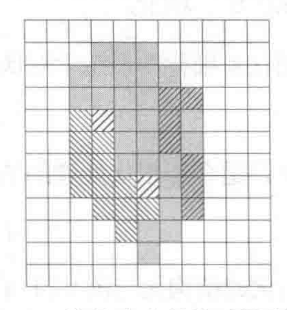


图 9.20　将凸壳生长算法限制到原始点集沿垂直方向和水平方向的最大尺寸的结果

9.5.5　细化

结构元 B 对集合 A 的细化可表示为 $A \otimes B$，它可以根据击中或击不中变换来定义：

$$A \otimes B = A - (A \circledast B) = A \cap (A \circledast B)^c \tag{9.5-6}$$

如前一节那样，我们仅对与结构元的模式匹配感兴趣，所以在击中或击不中变换中没有背景运算。对称地细化 A 的一种更有用表达方式是以结构元序列为基础的：

$$\{B\} = \{B^1, B^2, B^3, \cdots, B^n\} \tag{9.5-7}$$

式中 B^i 是 B^{i-1} 旋转后的形式。使用这一概念，我们现在可以使用一个结构元序列将细化定义为

$$A \otimes \{B\} = ((\cdots((A \otimes B^1) \otimes B^2) \cdots) \otimes B^n) \tag{9.5-8}$$

这种处理是 A 被 B^1 细化一次，然后，得到的结果被 B^2 细化一次，如此进行下去，直到 A 被 B^n 细化一次。整个过程不断重复，直到得到的结果不再发生变化。每次单独的细化均使用式(9.5-6)来执行。

图 9.21(a)显示了一组通常用于细化的结构元，图 9.21(b)显示了将要使用刚才讨论的过程来细化的集合 A。图 9.21(c)显示了 A 被 B^1 细化一次后得到的结果。图 9.21(d)到(k)显示了用其他结构元细化多次的结果。使用 B^6 进行第二次细化后得到了收敛的结果。图 9.21(l)显示了细化的结果。最后，图 9.21(m)显示了转换为 m 连通的细化集合(见 2.5.2 节)，以消除多重路径。

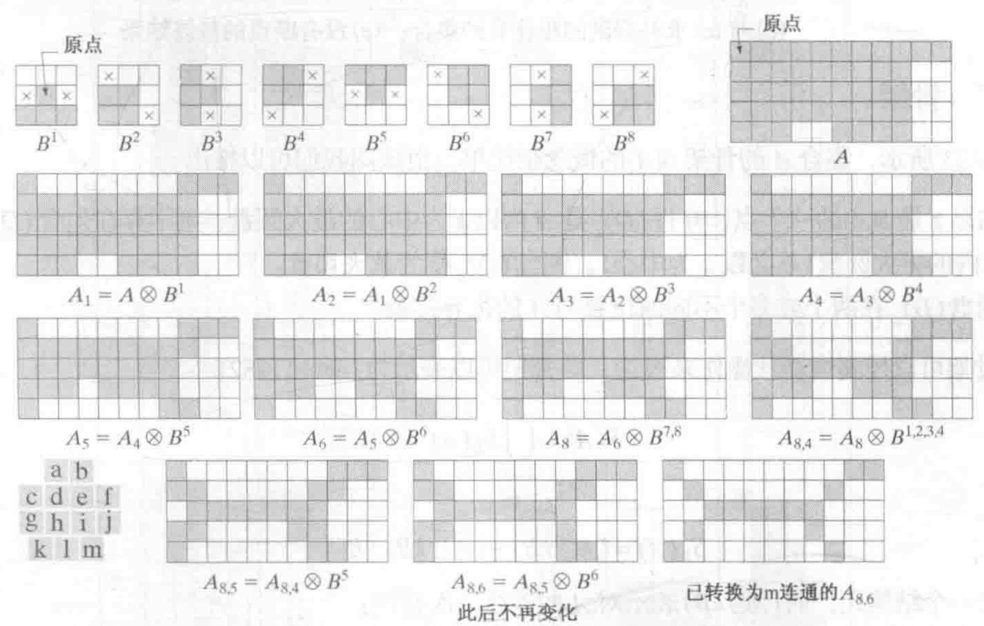


图 9.21　(a)用于细化的经旋转后的结构元序列；(b)集合 A；(c)使用第一个结构元细化得到的结果；(d)~(i)使用接下来的 7 个结构元细化得到的结果(使用第 7 个和第 8 个结构元细化得到的结果间没有变化)；(j)再次使用前 4 个结构元细化得到的结果；(l)收敛后的结果；(m)转换为 m 连通的结果

671 ~ 672

9.5.6 粗化

粗化是细化的形态学对偶，定义如下：

$$A \bullet B = A \cup (A \circledast B) \tag{9.5-9}$$

式中 B 是适合于粗化处理的结构元。与细化一样，粗化处理也可以定义为一个系列操作：

$$A \odot \{B\} = ((\cdots((A \odot B^1) \odot B^2)\cdots) \odot B^n) \tag{9.5-10}$$

用于粗化的结构元与图 9.21(a)所示的结构元具有相同的形式，但所有 1 和 0 要互换。然而，针对粗化的分离算法在实际中很少用到，取而代之的过程是先对问题中集合的背景进行细化，而后对结果求补集。换句话说，为粗化集合 A，我们先形成 $C = A^c$，而后细化 C，然后再求 C^c。图 9.22 说明了这一过程。

由于依赖于 A 的性质，这个过程可能会产生某些断点，如图 9.22(d)所示。因此，通过这种方法的粗化处理通常会跟随一个消除断点的后处理。注意，根据图 9.22(c)可知，细化后的背景形成了粗化处理的一个边界。这一有用的特性在使用式(9.5-10)粗化的直接执行中并不存在，并且它是使用背景细化来实现粗化的主要原因之一。

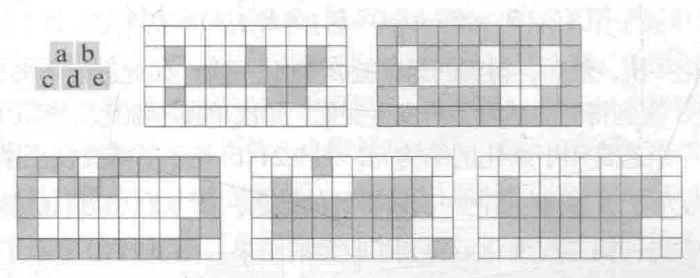


图 9.22 (a)集合 A；(b)A 的补集；(c)对 A 的补集进行细化得到的结果；(d)对(c)求补得到的粗化后的集合；(e)没有断点的最终结果

9.5.7 骨架

如图 9.23 所示，集合 A 的骨架 $S(A)$的概念很简单。由该图我们可以推出：

(a) 如果 z 是 $S(A)$的一个点，并且$(D)_z$ 是 A 内以 z 为中心的最大圆盘，则不存在包含$(D)_z$ 且位于 A 内的更大圆盘(不必以 z 为中心)。圆盘$(D)_z$ 称为最大圆盘。

(b) 圆盘$(D)_z$ 在两个或多个不同的位置与 A 的边界接触。

A 的骨架可以用腐蚀和开操作来表达，即骨架可以表示为(Serra[1982])

$$S(A) = \bigcup_{k=0}^{K} S_k(A) \tag{9.5-11}$$

其中，

$$S_k(A) = (A \ominus kB) - (A \ominus kB) \circ B \tag{9.5-12}$$

式中，B 是一个结构元，而$(A \ominus kB)$表示对 A 的连续 k 次腐蚀：

$$(A \ominus kB) = ((\cdots(A \ominus B) \ominus B) \ominus \cdots) \ominus B) \tag{9.5-13}$$

672 ~ 674

K 是 A 被腐蚀为空集前的最后一次迭代步骤。换句话说，

$$K = \max\{k \mid (A \ominus kB) \neq \varnothing\} \tag{9.5-14}$$

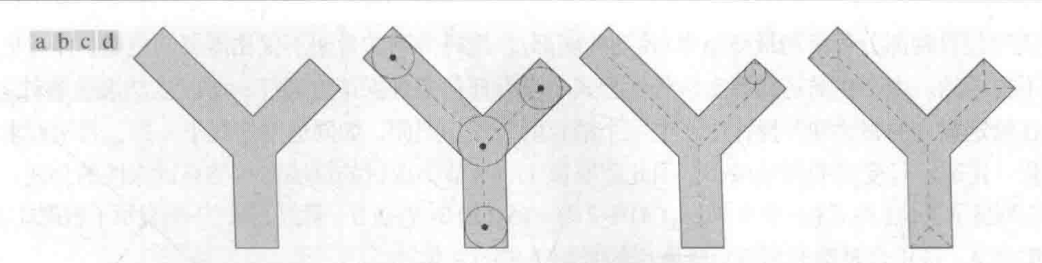


图 9.23　(a) 集合 A；(b) 中心在 A 的骨架上的最大圆盘的不同位置；(c) 位于 A 的骨架的不同线段上的另一个最大圆盘；(d) 完整的骨架

式 (9.5-11) 和式 (9.5-12) 给出的公式表明，$S(A)$ 可以作为骨架子集 $S_k(A)$ 的并集来得到。此外，还可证明使用下式可由这些子集来重建 A：

$$A = \bigcup_{k=0}^{K} (S_k(A) \oplus kB) \tag{9.5-15}$$

式中 $(S_k(A) \oplus kB)$ 表示对 $S_k(A)$ 的 k 次连续膨胀，即

$$(S_k(A) \oplus kB) = ((\cdots((S_k(A) \oplus B) \oplus B) \oplus \cdots) \oplus B) \tag{9.5-16}$$

例 9.8　计算简单图形的骨架。

图 9.24 说明了刚才讨论的概念。第一列显示了原始集合 (顶部) 和使用结构元 B 的两次腐蚀。注意，对 A 再进行一次腐蚀将产生空集，因此，在这种情况下 $K = 2$。第二列显示了使用 B 对第一列中的集合进行的开操作。结合图 9.8 中讨论的开操作的拟合特性，这些结果很容易解释。第三列只包含第一列和第二列间的差集。

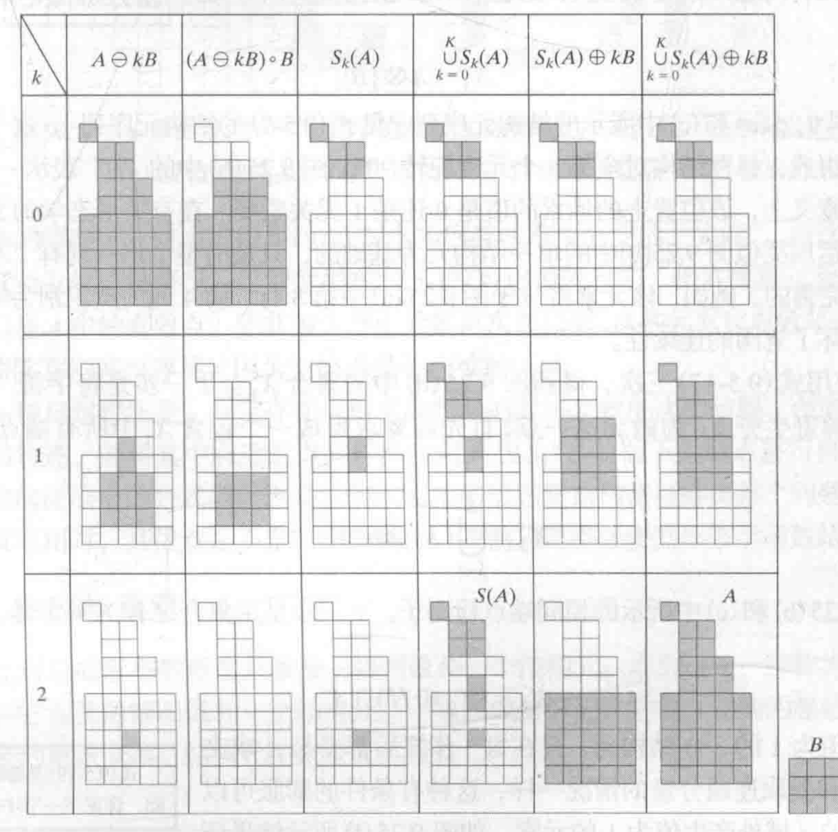


图 9.24　式 (9.5-11) 到式 (9.5-15) 的执行。原始集合位于左上角，其形态学骨架位于第四列的底部。第六列底部为重建的集合

674 ~ 675

第四列包含两部分骨架和最终结果(该列的底部)。最终得到的骨架不仅比需要的更粗，而且更重要的是它是不连通的。由于在前述的形态学骨架公式中没有任何条件保证连通性，所以该结果是意料之中的。形态学在给定集合的腐蚀和开操作中产生一个精致的公式。然而，如同通常情况下一样，骨架要求最大限度的细化、连通，且受到的腐蚀最小，因此需要像 11.1.7 节中探讨的算法这一类有试探性的公式。

第五列显示了 $S_0(A), S_1(A)\oplus B$ 和 $(S_2(A)\oplus 2B)=(S_2(A)\oplus B)\oplus B$ 。最终，最后一列显示了根据式(9.5-15)重建的集合 A，该集合是第五列所示膨胀后的骨架子集的并集。

9.5.8 裁剪

裁剪方法本质上是对细化和骨架算法的补充，因为这些过程会保留某些寄生成分，因而需要用后处理来清除这些寄生成分。我们首先探讨裁剪问题，然后以前面小节中介绍的内容为基础探讨一种形态学上的解决方法。因此，我们将说明如何联合使用已讨论的几种技术来解决这一问题。

在手写字符的自动识别中，通常使用的方法是分析每个字符的骨架形状。这些骨架经常带有许多"毛刺"(寄生成分)。毛刺是在腐蚀过程中由组成字符的笔画的不均匀性造成的。为处理这一问题，我们将探讨一种形态学技术，先假定寄生成分的长度不超过指定的像素数。

图 9.25(a)显示了手写字符"a"的骨架。该字符最左侧的寄生成分就是我们要消除的部分。解决方案是通常不断删除寄生分支的终点来抑制寄生分支。当然，这也会使得字符的其他分支变短(或删除)，但在缺乏其他结构信息的情况下，这个例子中的假设是任何具有三个或小于三个像素长度的分支都将被删除。使用一系列仅设计用来检测端点的结构元，对输入集合 A 进行细化可以得到期望的结果。也就是说，令

> 我们可将端点定义为一个大小为 3×3 的区域的中心点，该区域满足图 9.25(b)或(c)中的任何排列。

$$X_1 = A\otimes\{B\} \tag{9.5-17}$$

式中，$\{B\}$表示图 9.25(b)和(c)中所示的结构元序列[见式(9.5-7)的结构元序列]。这个结构元序列由两种不同的结构组成，每种结构对全部 8 个元素旋转 90°。图 9.25(b)中的"×"表示一个"不必考虑"的条件，在这种意义上，该位置上的像素的值是 0 还是 1 无关紧要。在有关形态学的文献中，报告的许多结果都是以运用类似图 9.25(b)中的单一结构元为基础的，只是沿整个第一列有"不必考虑"的条件而已。这是不完善的。例如，该元素将位于图 9.25(a)中第 8 行、第 4 列的点识别为端点，因此将该点消除，从而破坏了笔画的连续性。

连续对 A 应用式(9.5-17)三次，得到图 9.25(d)中的集合 X_1。下一步是将字符"复原"为原来的形状，但要去掉寄生分支。为做到这一点，首先需要求形成一个包含 X_1 中所有端点的集合 X_2[见图 9.25(e)]：

$$X_2 = \bigcup_{k=1}^{8}(X_1 \circledast B^k) \tag{9.5-18}$$

式中，B^k 是图 9.25(b)和(c)中所示的相同端点检测子。下一步是用集合 A 作为限定器，对端点进行三次膨胀：

$$X_3 = (X_2\oplus H)\cap A \tag{9.5-19}$$

式中，H 是元素值为 1 的 3×3 结构元，且在每一步之后都要与 A 求交集。如区域填充和提取连通分量的情况一样，这种有条件的膨胀可以防止在我们关注的区域外产生值为 1 的元素，如图 9.25(f)所示结果证明的那样。最后，X_3 和 X_1 的并集就是我们想要的结果：

> 式(9.5-19)是膨胀形态学重建的基础，详见下一节中的解释。



$$X_4 = X_1 \cup X_3 \tag{9.5-20}$$

如图 9.25(g) 所示。

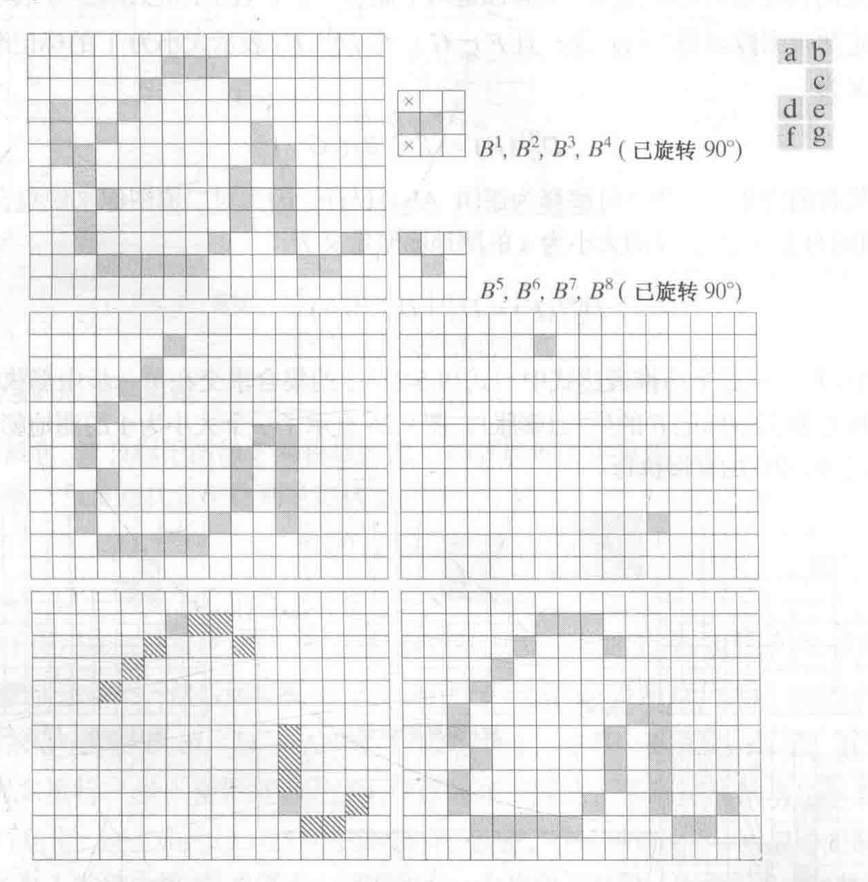


图 9.25　(a) 原图像；(b)～(c) 用于检测端点的结构元；(d) 经过三轮细化后的结果；(e) 图 (d) 的端点；(f) 以图 (a) 为条件的端点膨胀；(g) 修剪后的图像

在更复杂的情况下，使用式 (9.5-19) 有时会捡回某些寄生分支的"尖端"。当这些分支的端点离骨架很近时，就会发生这种情况。尽管使用式 (9.5-17) 可以消除它们，但在膨胀过程中会再次捡回这些点，因为它们是 A 中的有效点。除非整个寄生元素再次被捡回（这些元素比有效笔画短的情况很少），否则检测并消除它们就很容易，因为它们都是不连续的区域。

此时的一种自然想法是，一定存在一种更为容易的方法来解决这个问题。例如，我们可以保留所有被删除点的轨迹，并将其中合适的点与那些应用式 (9.5-17) 后留下的端点进行再连接。这样的选择是可行的，但前面给出的公式的优点是，使用简单的形态学结构即可解决整个问题。在实际情况下，当这一组工具可用时，其优点是不必写新的算法。我们只需将必要的形态学函数结合到操作序列中。

9.5.9　形态学重建

迄今为止讨论的形态学概念只涉及一幅图像和一个结构元。本节讨论一种称为形态学重建的强大形态学变换，它涉及两幅图像和一个结构元。一幅图像是标记，它包含变换的起始点，另一幅图像是模板，它用来约束该变换。结构元用来定义连接性[①]。

① 在关于形态学重建的许多文献中，默认情形下结构元假设是各向同性的，且一般称为初级各向同性结构元。在本章的上下文中，这样的一个结构元的例子是其原点在中心处、元素为 1 且大小为 3×3 的阵列。

测地膨胀和腐蚀

形态学重建的核心是测地膨胀和测地腐蚀这两个概念。令 F 表示标记图像，G 表示模板图像。在本讨论中，假定两幅图像都是二值图像，且 $F \subseteq G$。令 $D_G^{(1)}(F)$ 表示大小为 1 的标记图像关于模板的测地膨胀，定义为

$$D_G^{(1)}(F) = (F \oplus B) \cap G \tag{9.5-21}$$

其中，$\cap$ 表示集合的交集［这里 $\cap$ 可解释为逻辑 AND(与)，因为对二值图像来说集合的交集和逻辑 AND 操作是相同的］。F 关于 G 的大小为 n 的测地膨胀定义为

$$D_G^{(n)}(F) = D_G^{(1)}\left[D_G^{(n-1)}(F)\right] \tag{9.5-22}$$

式中，$D_G^{(0)}(F) = F$。在这个递推表达式中，式(9.5-21)中的集合求交在每一步中都执行①。注意，交集算子保证模板 G 将限制标记 F 的生长(膨胀)。图 9.26 显示了一个大小为 1 的测地膨胀的简单例子。图中的步骤是式(9.5-21)的直接执行。

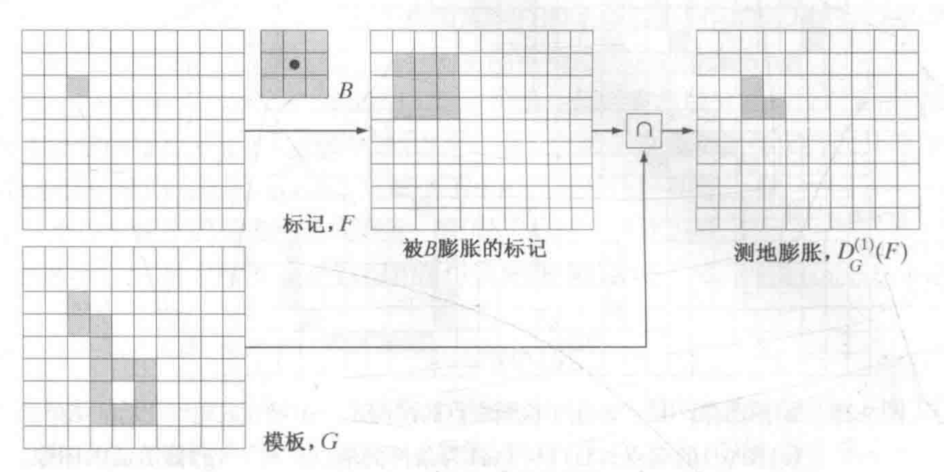


图 9.26 测地膨胀的说明

类似地，标记 F 关于模板 G 的大小为 1 的测地腐蚀定义为

$$E_G^{(1)}(F) = (F \ominus B) \cup G \tag{9.5-23}$$

式中，$\cup$ 表示集合的并集(或 OR 操作)。F 关于 G 的大小为 n 的测地腐蚀定义为

$$E_G^{(n)}(F) = E_G^{(1)}\left[E_G^{(n-1)}(F)\right] \tag{9.5-24}$$

式中，$E_G^{(0)}(F) = F$。式(9.5-23)中的并集操作在每个迭代步骤中执行，并保证一幅图像的测地腐蚀仍然大于等于其模板图像。如从式(9.5-21)和式(9.5-23)预期的那样，测地膨胀和测地腐蚀关于集合的补集对偶(见习题 9.29)。图 9.27 显示了一个大小为 1 的测地腐蚀的简单例子。图中的步骤是直接执行式(9.5-23)。

有限数量图像的测地膨胀和腐蚀经过有限数量的迭代步骤后总会收敛，因为标记图像的扩散或收缩受模板约束。


678~680

① 尽管使用递归公式来开发形态学重建方法更为直观(就像此处所做的那样)，但它们的实际实现通常基于计算上更有效率的算法(例如，见 Vincent[1993]和 Soille[2003])。本节中所有基于图像的例子均是使用这样的算法生成的。

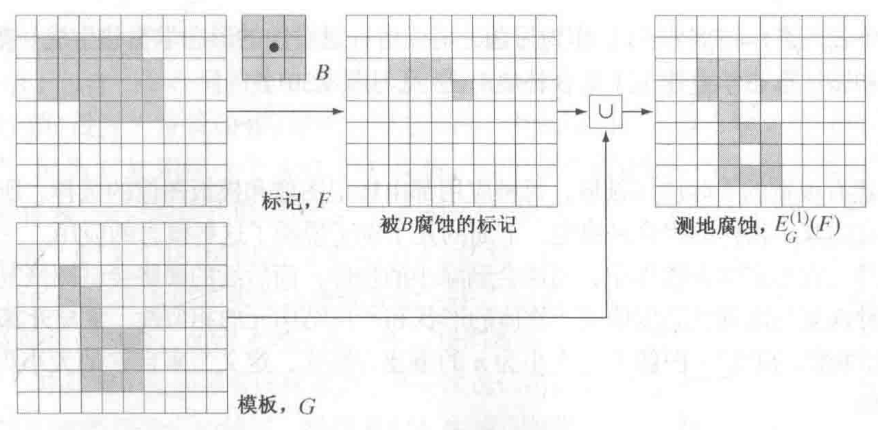


图 9.27　测地腐蚀的说明

用膨胀和腐蚀的形态学重建

基于前面的概念，来自标记图像 F 对模板图像 G 的膨胀形态学重建表示为 $R_G^D(F)$，它被定义为 F 关于 G 的测地膨胀，反复迭代直至达到稳定状态；即

$$R_G^D(F) = D_G^{(k)}(F) \tag{9.5-25}$$

迭代 k 次，直至 $D_G^{(k)}(F) = D_G^{(k+1)}(F)$。

图9.28说明了使用膨胀的重建。图9.28(a)继续在图9.26中开始的处理；也就是说，得到 $D_G^{(1)}(F)$ 后，重建的下一步是膨胀该结果，然后用模板 G 与其相“与”(AND)得到 $D_G^{(2)}(F)$，如图 9.28(b)所示。$D_G^{(2)}(F)$ 的膨胀结果与模板 G 相“与”(AND)得到 $D_G^{(3)}(F)$，等等。重复这一过程，直至达到稳定。如果我们对该例多执行一步，会发现 $D_G^{(5)}(F) = D_G^{(6)}(F)$，因此，采用膨胀的形态学重建的图像由 $D_G^D(F) = D_G^{(5)}(F)$ 给出，这正如式(9.5-25)指出的那样。注意，在这种情况下，重建的图像与模板相同，因为 F 包含了值为 1 的单个像素(类似于一幅图像与一个冲激的卷积，该卷积简单地复制冲激位置处的图像，如在 3.4.2 节解释的那样)。

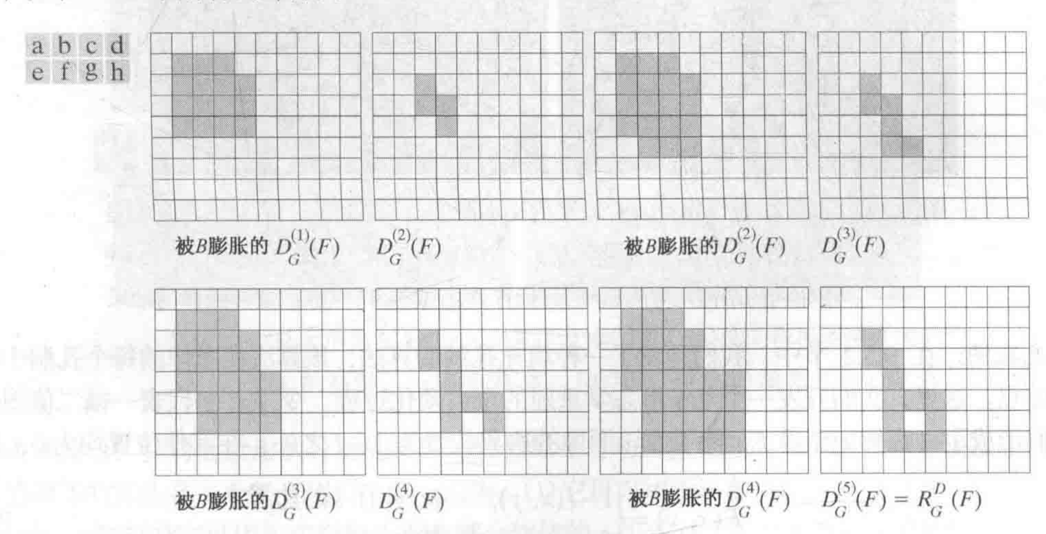


图 9.28　采用膨胀的形态学重建的说明。F, G, B 和 $D_G^{(1)}(F)$ 来自图 9.26

依照类似的方法，模板图像 G 对标记图像 F 的腐蚀的形态学重建表示为 $R_G^E(F)$，它被定义为 F 关于 G 的测地腐蚀，反复迭代直至达到稳定状态；即

680

$$R_G^E(F) = E_G^{(k)}(F) \tag{9.5-26}$$

迭代 k 次，直至 $E_G^{(k)}(F)=E_G^{(k+1)}(F)$。作为习题，请读者针对腐蚀的形态学重建生成一幅类似于图9.28的图形。膨胀和腐蚀形态学重建关于集合补集对偶(见习题 9.30)。

应用实例

形态学重建有很宽的实际应用领域，每种应用都由标记图像和模板图像的选择、所用的结构元及前面讨论中定义的基本操作的组合来决定。下面的几个例子说明了这些概念的应用。

重建开操作：在形态学开操作中，腐蚀会删除小的物体，而后续的膨胀会试图恢复遗留物体的形状。然而，这种恢复的准确性高度依赖于物体的形状和所用结构元的相似性。重建开操作可正确地恢复腐蚀后所保留物体的形状。图像 F 的大小为 n 的重建开操作，定义为来自 F 的大小为 n 的腐蚀的 F 的膨胀重建；即

$$O_R^{(n)}(F)=R_F^D[(F\ominus nB)] \tag{9.5-27}$$

式中，$(F\ominus nB)$ 表示 B 对 F 的 n 次腐蚀，如 9.5.7 节中解释的那样。注意，在该应用中，F 被用作模板。对于重建闭操作，可写出一个类似的表达式(见表 9.1)。

图 9.29 显示了重建开操作的一个例子。在这一说明中，我们的兴趣是从图 9.29(a) 中提取垂直笔画的长字符。重建开操作至少要求一次腐蚀，因此我们首先执行这一步骤。图 9.29(b) 显示了图 9.29(a) 经腐蚀后的结果，该腐蚀采用长度与长字符的(51 个像素)平均高度成正比和宽度为 1 个像素的结构元。为了比较，我们用相同的结构元计算了开操作。图 9.29(c) 显示了结果。最后，图 9.29(d) 是用式(9.5-27)中给出的 F［即 $O_R^{(1)}(F)$］的重建开操作(大小为 1)。该结果表明，包含垂直笔画的长字符被准确地恢复，而所有其他字符则被去除。

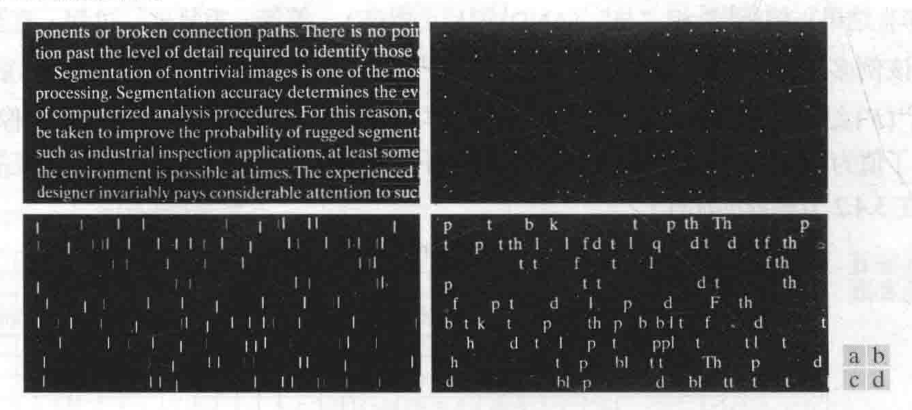


图 9.29 (a)大小为 918×2018 像素的文本图像。长的字符的近似平均高度是 51 个像素；(b)使用大小为 51×1 像素的结构元对图(a)的腐蚀；(c)使用相同的结构元对图(a)的开操作，用做参考；(d)重建开操作的结果

填充孔洞：在 9.5.2 节中，我们开发了一种填充孔洞的算法，该算法在图像的每个孔洞中都已知一个起始点。这里，我们开发一个基于形态学重建的全自动化过程。令 $I(x, y)$代表一幅二值图像，并假定我们形成了一幅标记图像 F，除了在该图像的边界位置为 $1-I$ 之外，在其他位置均为 0，即

$$F(x,y)=\begin{cases}1-I(x,y), & (x,y)\text{ 在 } I \text{ 的边界上}\\ 0, & \text{其他}\end{cases} \tag{9.5-28}$$

则

$$H=\left[R_{I^c}^D(F)\right]^c \tag{9.5-29}$$

是一幅等于 I 且所有孔洞都被填充的二值图像。

680 ~ 682

让我们考虑式(9.5-29)中的各个分量，以了解该表达式事实上是如何填充图像中的所有孔洞的。图 9.30(a)显示了包含一个孔洞的简单图像 I，图 9.30(b)显示了它的补集。注意，因为 I 的补集将所有的前景像素(1 值)设置为背景(0 值)像素，反之亦然，因此该操作效果上建立了围绕孔洞的一面元素为 0 的“墙”。因为 I^c 被用作一个 AND(与)模板，这里我们所做的一切都是在该过程的迭代期间保护所有的前景像素(包括围绕孔洞的墙)。图 9.30(c)是根据式(9.5-28)形成的阵列 F，图 9.30(d)是使用所有元素都为 1 的 3×3 SE(结构元)膨胀后的 F。注意，标记 F 有一个元素为 1 的边缘(除 I 为 1 的位置之外)，故这些标记点 F 的膨胀在边缘处开始，并且向内处理。图 9.30(e)显示了使用 I^c 作为模板对 F 进行测地膨胀后的结果。如刚才指出的那样，我们看到，在这个结果中，对应于 I 的前景像素的所有位置都是 0，并且对于孔洞像素同样如此。另一次迭代将得到相同的结果，如式(9.5-29)要求的那样，求补后给出图 9.30(f)中所示的结果。操作 $H \cap I^c$ 生成一幅图像，该图像中对应于 I 中孔洞的位置像素值为 1，如图 9.30(g)所示。

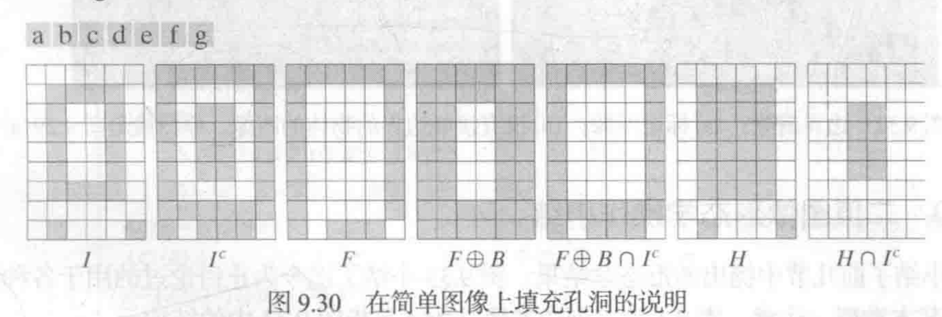


图 9.30　在简单图像上填充孔洞的说明

图 9.31 显示了一个更实际的例子。图 9.31(b)显示了图 9.31(a)中文本图像的补集，图 9.31(c)是由式(9.5-28)产生的标记图像 F。除了对应于原图像边界上的 1 的位置外，该图像有一个元素为 1 的边界。最后，图 9.31(d)显示了所有孔洞均被填充后的图像。

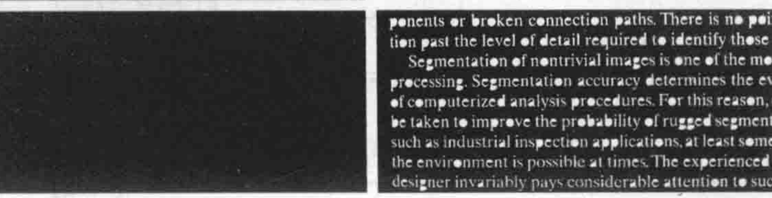


图 9.31　(a)大小为 918×2018 像素的文本图像；(b)作为模板图像使用的(a)的补集；(c)标记图像；(d)用式(9.5-29)填充孔洞的结果

边界清除：针对后续形状分析的图像物体提取是自动图像处理的基本任务。删除接触(即连接到)边界的物体的算法是一个很有用的工具，因为：(1)它可以用于屏蔽图像，以便对进一步处理只保留完整的物体，或者(2)它可用作在视野中存在部分物体的一个信号。作为本节介绍的概念的最后说明，我们现在提出一个基于形态学重建的边界清除过程。在这一应用中，我们用原图像作为模板，并使用下面的标记图像：

$$F(x, y) = \begin{cases} I(x, y), & (x, y)\text{位于 } I \text{ 的边界上} \\ 0, & \text{其他} \end{cases} \tag{9.5-30}$$

682 ~ 683

边界清除算法首先计算形态学重建 $R_I^D(F)$（简单地提取接触到边界的物体），然后计算差

$$X = I - R_I^D(F) \tag{9.5-31}$$

以得到一幅其中没有接触边界的物体的图像 X。

作为一个例子，再次考虑文本图像。图 9.32(a) 显示了使用所有元素均为 1 的 3×3 的结构元得到的重建 $R_I^D(F)$（注意右侧接触边界的物体），图 9.32(b) 显示了使用式(9.5-31)计算得到的图像 X。如果手边的任务是自动字符识别，则拥有一幅其中没有接触边界的字符的图像最有用，因为这样可以避免不得不识别部分字符的问题(充其量也只是一个困难的任务)。

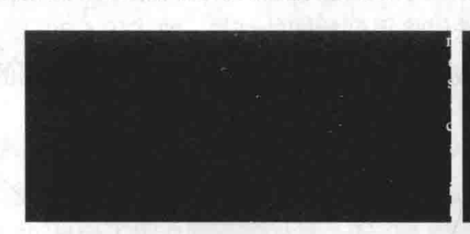


图 9.32 边界清除：(a) 标记图像；(b) 没有接触边界的物体的图像。原图像是图 9.29(a)

9.5.10 二值图像形态学操作小结

表 9.1 小结了前几节中提出的形态学结果。图 9.33 小结了迄今为止讨论过的用于各种形态学处理的结构元的基本类型。注意，表 9.1 第三列中的罗马数字是指图 9.33 中的结构元。

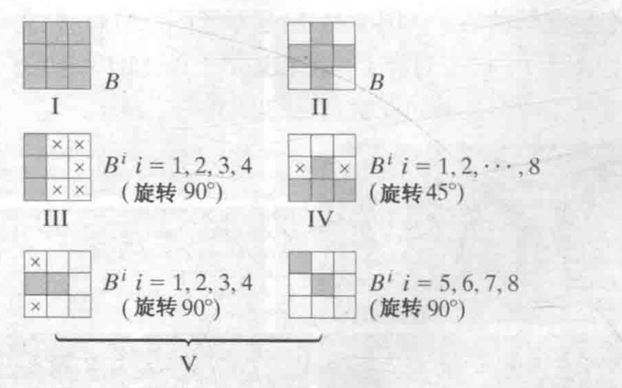


图 9.33 用于二值形态学的 5 个基本结构元。每个结构元的原点均位于其中心处。×项表示“不考虑”的值

表 9.1 形态学操作及其性质小结

操 作	公 式	注 释 (罗马数字指图 9.33 中的结构元)
平移	$(B)_z = \{w \mid w = b + z,\ b \in B\}$	将 B 的原点平移到点 z
反射	$\hat{B}_z = \{w \mid w = -b,\ b \in B\}$	关于集合 B 的原点映射该集合的所有元素
求补集	$A^c = \{w \mid w \notin A\}$	不属于 A 的点的集合
求差集	$A - B = \{w \mid w \in A,\ w \notin B\}$ $= A \cap B^c$	属于 A 但不属于 B 的点的集合
膨胀	$A \oplus B = \{z \mid (\hat{B}_z) \cap A \neq \varnothing\}$	“扩展” A 的边界(I)
腐蚀	$A \ominus B = \{z \mid (B)_z \subseteq A\}$	“收缩” A 的边界(I)
开操作	$A \circ B = (A \ominus B) \oplus B$	平滑轮廓，切断狭窄区域，并消除小的孤岛和尖刺(I)
闭操作	$A \bullet B = (A \oplus B) \ominus B$	平滑轮廓，融合狭窄间断和细长沟壑，并消除小的孔洞(I)

683
~
685

（续表）

操　作	公　式	说　明 （罗马数字是指图9.33中的结构元）
击中或击不中变换	$A\circledast B=(A\ominus B_1)\cap(A^c\ominus B_2)$ $=(A\ominus B_1)-(A\oplus \hat{B}_2)$	点(坐标)的集合，在这样的点处，同时，B_1在A中可找到一个匹配(击中)，B_2在A^c中可找到一个匹配
边界提取	$\beta(A)=A-(A\ominus B)$	在集合A的边界上的点的集合(I)
孔洞填充	$X_k=(X_{k-1}\oplus B)\cap A^c$; $k=1,2,3,\cdots$	填充A中的孔洞；X_0是每个孔洞处元素为1而其他位置元素为0的阵列(II)
连通分量	$X_k=(X_{k-1}\oplus B)\cap A$; $k=1,2,3,\cdots$	寻找A中的连通分量；X_0是每个连通分量中元素为1而其他位置元素为0的阵列(I)
凸壳	$X_k^i=(X_{k-1}^i\circledast B^i)\cup A$; $i=1,2,3,4$; $k=1,2,3,\cdots$; $X_0^i=A$; $D^i=X_{\text{conv}}^i$	寻找集合A的凸壳$C(A)$，其中conv表示$X_k^i=X_{k-1}^i$ (III)意义上的收敛，
细化	$A\otimes B=A-(A\circledast B)$ $=A\cap(A\circledast B)^c$ $A\otimes\{B\}=$ $((\cdots((A\otimes B^1)\otimes B^2)\cdots)\otimes B^n)$ $\{B\}=\{B^1,B^2,B^3,\cdots,B^n\}$	细化集合A。前两个公式给出细化的基本定义。后两个公式表示使用一个结构元序列的细化。实际中通常使用这种方法(IV)
粗化	$A\odot B=A\cup(A\circledast B)$ $A\odot\{B\}=$ $((\cdots(A\odot B^1)\odot B^2\cdots)\odot B^n)$	粗化集合A(见前面关于结构元序列的说明)。使用IV，但把0和1颠倒
骨架	$S(A)=\bigcup_{k=0}^{K}S_k(A)$ $S_k(A)=\{(A\ominus kB)$ $-[(A\ominus kB)\circ B]\}$ A的重建： $A=\bigcup_{k=0}^{K}(S_k(A)\oplus kB)$	寻找集合A的骨架$S(A)$。最后一个公式指出A可以由其骨架子集$S_k(A)$重建。在所有3个公式中，K是集合A被腐蚀为空集时的迭代次数。符号$(A\ominus kB)$表示B对A连续腐蚀的第k次迭代(I)
裁剪	$X_1=A\otimes\{B\}$ $X_2=\bigcup_{k=1}^{8}(X_1\circledast B^k)$ $X_3=(X_2\oplus H)\cap A$ $X_4=X_1\cup X_3$	X_4是裁剪集合A后的结果。必须指定使用第一个公式来得到X_1的次数。结构元V用于前两个公式。在第三个公式中，H表示结构元I
大小为1的测地膨胀	$D_G^{(1)}(F)=(F\oplus B)\cap G$	F和G分别称为标记图像和模板图像
大小为n的测地膨胀	$D_G^{(n)}(F)=D_G^{(1)}\left[D_G^{(n-1)}(F)\right]$; $D_G^{(0)}(F)=F$	
大小为1的测地腐蚀	$E_G^{(1)}(F)=(F\ominus B)\cup G$	
大小为n的测地腐蚀	$E_G^{(n)}(F)=E_G^{(1)}\left[E_G^{(n-1)}(F)\right]$ $E_G^{(0)}(F)=F$	
膨胀形态学重建	$R_G^D(F)=D_G^{(k)}(F)$	k满足$D_G^{(k)}(F)=D_G^{(k+1)}(F)$
腐蚀形态学重建	$R_G^E(F)=E_G^{(k)}(F)$	k满足$E_G^{(k)}(F)=E_G^{(k+1)}(F)$
重建开操作	$O_R^{(n)}(F)=R_F^D[(F\ominus nB)]$	$(F\ominus nB)$表示B对F的n次腐蚀
重建闭操作	$C_R^{(n)}(F)=R_F^E[(F\oplus nB)]$	$(F\oplus nB)$表示B对F的n次膨胀
孔洞填充	$H=\left[R_{I^c}^D(F)\right]^c$	H等于输入图像I，但具有所有被填充的孔洞。标记图像F的定义见式(9.5-28)
边界清除	$X=I-R_I^D(F)$	X等于输入图像I，但带有所有被删除接触(连接到)边界的物体。标记图像F的定义见式(9.5-30)

685
~
686

9.6 灰度级形态学

在这一节中，我们将把膨胀、腐蚀、开操作和闭操作的基本操作扩展到灰度级图像。然后，我们将使用这些操作来探讨几个基本的灰度级形态学算法。

通过下面的讨论，我们将处理形如 $f(x, y)$ 和 $b(x, y)$ 的数字函数，其中 $f(x, y)$ 是一幅灰度级图像，而 $b(x, y)$ 是一个结构元。假设这些函数是 2.4.2 节中介绍过的离散函数。也就是说，如果 Z 表示实整数集合，则坐标 (x, y) 是来自笛卡儿积 Z^2 的整数，且 f 和 b 是对每个 (x, y) 坐标对赋以灰度值(来自实数集合 R 的一个实数)的函数。如果灰度级也是整数，则用 Z 代替 R。

灰度级形态学中的结构元所执行的基本功能与二值形态学中对应的功能相同：它们作为一个"探测器"以明确的特性检验一幅给定的图像。灰度级形态学中的结构元分为两类：非平坦的和平坦的结构元。图 9.34 显示了每类的一个例子。图 9.34(a) 以图像形式显示了一个半球状的灰度级结构元，图 9.34(c) 是一个通过其中心的水平灰度剖面。图9.34(b) 显示了一个圆盘形的平坦结构元，图9.34(d) 是其对应的灰度剖面(该剖面的形状解释了"平坦"一词的由来)。为清楚起见，图9.34 中的元素以连续量显示；它们的计算机实现是以数字近似为基础的(见图9.2 中最右侧的圆盘形 SE"结构元")。由于本节稍后讨论的一些困难，实际中灰度级结构元并不常用。最后，我们要提及的是，如在二值情况那样，必须清楚地确定结构元的原点。除非另外提及，本节中的所有例子都以高度为 1、对称的、平坦的结构元为基础，其原点位于中心处。灰度级形态学中结构元的反射的定义，如 9.1 节中的定义一样，且在下面的讨论中我们将它表示为 $\hat{b}(x, y) = b(-x, -y)$ 。

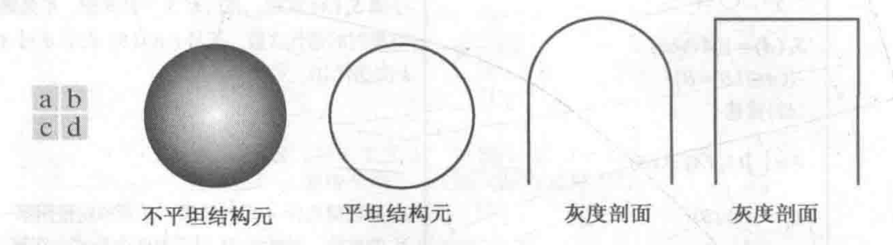


图 9.34 非平坦和平坦结构元，以及对应的通过其中心的水平灰度剖面。本节中的所有例子均基于平坦结构元

9.6.1 腐蚀和膨胀

当 b 的原点位于 (x, y) 处时，用一个平坦的结构元 b 在 (x, y) 处对图像 f 的腐蚀定义为图像 f 中与 b 重合区域的最小值。以公式的形式，结构元素 b 对一幅图像 f 在位置 (x, y) 处的腐蚀由下式给出：

$$[f \ominus b](x, y) = \min_{(s, t) \in b} \{f(x+s, y+t)\} \tag{9.6-1}$$

其中，用与 3.4.2 节讨论的相关过程类似的方式，x 和 y 通过所有要求的值增加，以便 b 的原点能访问 f 中的每个像素。也就是说，为寻求 b 对 f 的腐蚀，我们把结构元的原点放在图像的每个像素的位置。在任何位置的腐蚀由从包含在与 b 重合区域中的 f 的所有值中选取的最小值决定。例如，如果 b 是大小为 3×3 的方形结构元，则获得一点处的腐蚀要求寻找包含在由 b 定义的 3×3 区域中的 f 的 9 个值中，其原点所在的那个点的最小值。

类似地，当 $\hat{b}$ 的原点位于位置 (x, y) 处时，平坦结构元 b 在任何位置 (x, y) 处对图像 f 的膨胀，定义为图像 f 中与 $\hat{b}$ 重合区域的最大值，即

687 ~ 688

$$[f \oplus b](x, y) = \max_{(s,\, t)\in b} \{f(x-s,\, y-t)\} \tag{9.6-2}$$

其中，我们用到了早些时候说明过的 $\hat{b} = b(-x, -y)$ 这一事实。对该式的解释与前一段中的解释相同，但使用的是最大而非最小操作，并且要记住结构元关于其原点反射，我们用$(-s, -t)$考虑函数的自变量。这类似于空间卷积，就像 3.4.2 节中解释的那样。

例 9.9　灰度级腐蚀和膨胀的说明。

因为使用一个平坦 SE（结构元）的灰度级腐蚀计算图像 f 中与 b 重合的(x, y)的每个邻域的最小灰度值，因此通常我们希望被腐蚀的灰度级图像比原始图像暗，亮特征的尺寸（关于 SE 的尺寸）将会减小，而暗特征的尺寸将会增大。图 9.35(b) 显示了使用一个单位高度和半径为 2 像素的圆盘形结构元素对图9.35(a) 的腐蚀。刚才讨论的效果在腐蚀过的图像中清晰可见。例如，请注意小亮点亮度的降低程度，图 9.35(b) 中几乎已看不到这些小亮点，同时暗色特征则变浓。腐蚀过的图像中的一般背景比原图像的背景要稍暗一些。类似地，图9.35(c) 显示了使用相同 SE 膨胀后的结果。其效果与用腐蚀得到的效果相反，即亮特征变浓，而暗特征变淡。请特别注意图 9.35(a) 中左侧、中间、右侧和底部的较细黑色连线在图 9.35(c) 中几乎已看不见。黑点的尺寸已被减小，但与图9.35(b) 中被腐蚀的小白点完全不同，在膨胀后的图像中，小白点却清晰可见。其原因是与结构元的尺寸相比，黑点的尺寸原来就比白点大。最后，请注意膨胀后的图像中的背景比图 9.35(a) 中的背景稍亮。

非平坦结构元具有随定义域变化的灰度级。非平坦结构元 b_N 对图像 f 的腐蚀定义如下：

$$[f \ominus b_N](x, y) = \min_{(s,\, t)\in b_N} \{f(x+s,\, y+t) - b_N(s,\, t)\} \tag{9.6-3}$$

这里，我们实际上是从 f 中减去 b_N 的值来确定任意点处的腐蚀。与式(9.6-1)不同，这意味着使用非平坦结构元的腐蚀通常不受 f 值的限制，而这在解释结果时会存在问题。因为这一原因，实际中很少使用灰度级结构元；另外，与式(9.6-1)相比，为 b_N 选取有意义的元素和增加的计算负担也是一个问题。

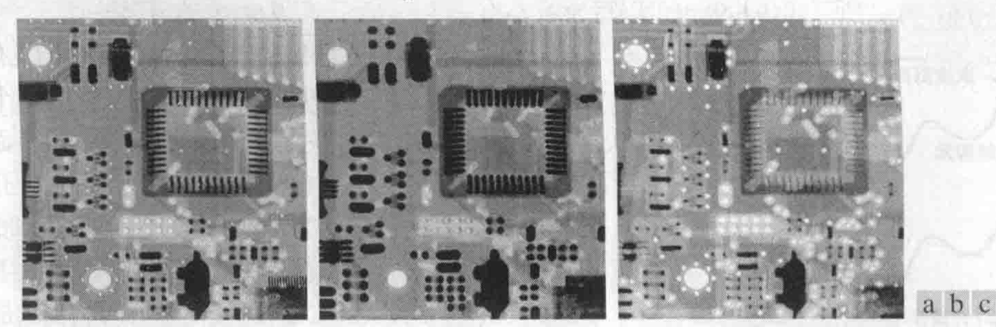


图 9.35　(a)大小为 448×425 像素的灰度级 X 射线图像；(b)使用半径为 2 个像素的圆盘形结构元对图像的腐蚀结果；(c)用相同结构元对图像的膨胀结果（原图像由 Lixi 公司提供）

采用类似的方式，使用非平坦结构元的膨胀定义如下：

$$[f \oplus b_N](x, y) = \max_{(s,\, t)\in b_N} \{f(x-s,\, y-t) + b_N(s,\, t)\} \tag{9.6-4}$$

前一段给出的说明同样适用于使用非平坦结构元的膨胀。当 b_N 的所有元素都是常数时（即结构元是平坦的），式(9.6-3)和式(9.6-4)就分别简化为式(9.6-1)和式(9.6-2)，其中一个标量常数等于该结构元的幅度。

如二值情况那样，腐蚀和膨胀是关于函数的补集和反射对偶的，即

$$(f \ominus b)^c(x, y) = (f^c \oplus \hat{b})(x, y)$$

688 ~ 689

式中，$f^c = -f(x,y)$ 且 $\hat{b} = b(-x,-y)$。相同的表达式对非平坦结构元也有效。为清楚起见，在下面的讨论中，我们将忽略所有函数的参量以简化表达式，在这种情况下，前一公式可写为

$$(f \ominus b)^c = (f^c \oplus \hat{b}) \tag{9.6-5}$$

类似地，

$$(f \oplus b)^c = (f^c \ominus \hat{b}) \tag{9.6-6}$$

自身的腐蚀和膨胀在灰度级图像处理中并不是特别有用。像对应的二值情形那样，当这些操作组合用来推导高级算法时，会变得强有力，如下节所示的材料那样。

> 尽管我们在本节剩余内容的例子中介绍的是平坦结构元，但讨论的概念同样适用于非平坦结构元。

9.6.2 开操作和闭操作

灰度级图像的开操作和闭操作的表达式，与二值图像的对应操作具有相同的形式。结构元 b 对图像 f 的开操作表示为 $f \circ b$，即

$$f \circ b = (f \ominus b) \oplus b \tag{9.6-7}$$

像之前那样，开操作先只用 b 对 f 做腐蚀，随后用 b 对所得结果做膨胀。类似地，b 对 f 的闭操作表示为 $f \bullet b$，即

$$f \bullet b = (f \oplus b) \ominus b \tag{9.6-8}$$

灰度级图像的开操作和闭操作关于函数的补集和结构元的反射是对偶的：

$$(f \bullet b)^c = f^c \circ \hat{b} \tag{9.6-9}$$

和

$$(f \circ b)^c = f^c \bullet \hat{b} \tag{9.6-10}$$

因为 $f^c = -f(x,y)$，所以式(9.6-9)也可以写为 $-(f \bullet b) = (-f \circ \hat{b})$，式(9.6-10)的情形类似。

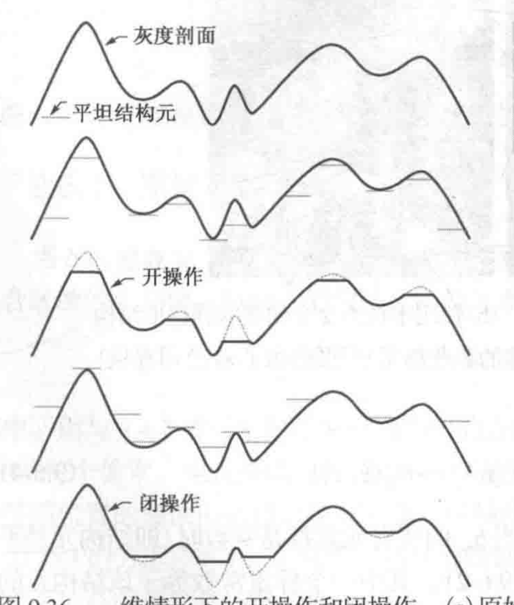


图 9.36 一维情形下的开操作和闭操作：(a)原始的一维信号；(b)从信号底部向上推动的平坦结构元素；(c)开操作；(d)沿信号顶部向下推动的平坦结构元素；(e)闭操作

图像的开操作和闭操作具有简单的几何解释。假设我们将一个图像函数 $f(x, y)$视为一个三维表面；也就是说，它的灰度值可解释为 xy 平面上的高度值，如图 2.18(a)所示。然后，b 对 f 的开操作可从几何角度解释为，从 f 的下表面向上推动结构元。在 b 的每个原点位置，开操作是当从 f 的下表面向上推动结构元时，b 的任何部分所达到的最高值。这样，完全开操作就是由 b 的原点访问 f 的每个坐标(x, y)所得到的所有值的集合。

图 9.36 以一维的形式说明了这个概念。假定图 9.36(a)中的曲线是沿图像中的一行的灰度剖面。图 9.36(b)显示了沿曲线底部向上推到不同位置的一个平坦结构元。图9.36(c)的实线是完全开操作。因为结构元太大而不能完全拟合曲线上部峰值的内侧，故峰值的顶部被开操作剪切掉了，剪切的量与结构元能够到达峰值的

> 有时，开操作和闭操作可由在一条曲线的下侧和上侧滚动一个圆来说明。在三维情形下，圆可用球体来替代，得到的过程称为滚球算法。

689 ~ 691

距离成比例。通常，开操作用于去除较小的明亮细节，而保持整体灰度级和较大的明亮特征相对不变。

图9.36(d)是闭操作的图形说明。很明显，结构元从曲线的顶部向下推动，同时平移到所有位置。如图9.36(e)所示，当结构元从曲线的上侧滑动时，通过寻找结构元的任何部分所能到达的最低点，即可构建闭操作。

灰度级开操作满足如下性质：

(a) $f \circ b \dashv f$ 。

(b) 若 $f_1 \dashv f_2$ ，则 $(f_1 \circ b) \dashv (f_2 \circ b)$ 。

(c) $(f \circ b) \circ b = (f \circ b)$ 。

符号 $e \dashv r$ 用来表示 e 的域是 r 的域的一个子集，且对于 e 的域中的任何(x, y)，有 $e(x, y) \leqslant r(x, y)$。类似地，闭操作满足如下性质：

(a) $f \dashv f \bullet b$ 。

(b) 若 $f_1 \dashv f_2$ ，则 $(f_1 \bullet b) \dashv (f_2 \bullet b)$ 。

(c) $(f \bullet b) \bullet b = (f \bullet b)$ 。

这些性质的用途类似于二值图像中的对应性质。

例 9.10　灰度级开操作和闭操作的说明。

图 9.37 将图 9.36 中说明的一维概念拓展到了二维情形。图 9.37(a)与我们在例 9.9 使用的图像相同，图 9.37(b)是使用单位高度和半径为 3 像素的圆盘形结构元得到的开操作结果。如预料的那样，所有亮特征的灰度都降低了，降低的程度取决于这些特征相对于结构元的尺寸。该图与图9.35(b)相比，我们看到，与腐蚀的结果不同，开操作对图像的暗特征影响可忽略不计，也不影响背景。类似地，图 9.37(c)显示了使用半径为 5 的圆盘形结构元得到的闭操作结果(小圆黑点比小白点大，因此要达到可与开操作相比的结果，需要更大的圆盘形结构元)。在这幅图像中，亮细节和背景相对来说未受影响，但削弱了暗特征，削弱的程度取决于这些特征相对于结构元的尺寸。

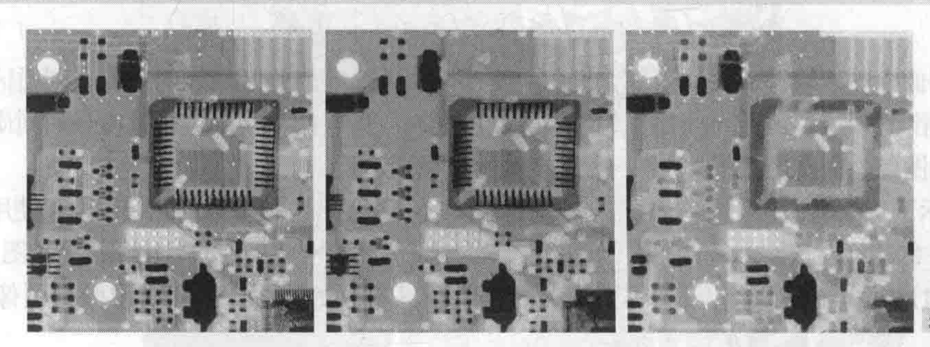


图 9.37　(a)一幅大小为 448×425 像素的灰度级 X 射线图像；(b)使用半径为 3 个像素的圆盘形结构元得到的开操作结果；(c)使用半径为 5 个像素的结构元得到的闭操作结果

9.6.3　一些基本的灰度级形态学算法

许多形态学技术是以迄今为止介绍的灰度级形态学概念为基础的。下面的讨论中我们将说明这些算法。

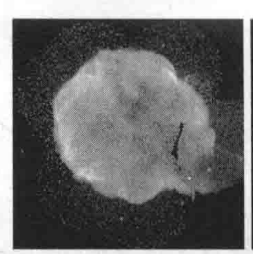
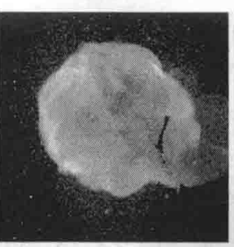
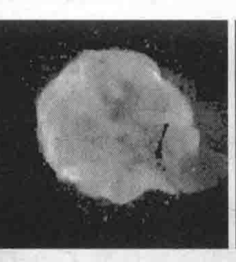
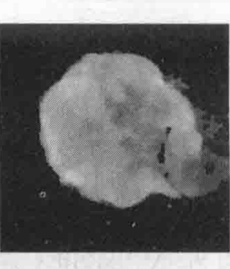
形态学平滑

因为开操作抑制比结构元小的亮细节，而闭操作抑制暗细节，所以它们常常以形态滤波的形式结合起来平滑图像和去除噪声。考虑图 9.38(a)，它显示了一幅使用 X 射线波段拍摄的天鹅星座

691～692

环超新星图像(关于该图像的细节请参阅图 1.7)。为当前讨论的目的，假设中心的亮区域是我们感兴趣的目标，而较小的分量是噪声。我们的目的是去除噪声。图9.38(b)显示了用一个半径为 2 的平坦圆盘(结构元)对原始图像进行开操作，再用相同尺寸的 SE(结构元)进行闭操作的结果。图 9.38(c)和(d)分别显示了使用半径为 3 和 5 的结构元进行相同操作得到的结果。如期望的那样，该图像序列显示了消除作为结构元尺寸的函数的小分量的改进。在最后的结果中，我们看到，感兴趣的目标已被提取出来。图像底部的噪声分量未被完全去除，原因在于它们的密度。

图9.38的结果是以对原始图像进行开操作，然后再进行闭操作为基础的。有时使用的一个过程是执行交替顺序滤波，在该过程中，对原始图像以开操作-闭操作顺序开始，但后续步骤对前一步骤的结果再执行开操作和闭操作。这种类型的滤波在自动图像分析中很有用，滤波过程中每一步的结果都针对一个指定的度量进行比较。通常，对于相同大小的结构元，与图9.38中说明的方法相比，这种方法会产生更为模糊的结果。

a b c d

图 9.38 (a)由 NASA 的哈勃望远镜在 X 射线波段拍摄的天鹅星座环超新星的 566×566 图像；(b)~(d)分别使用半径为 1，3 和 5 的圆盘形结构元对原图像执行开操作和闭操作顺序的结果(原图像由 NASA 提供)

形态学梯度

膨胀和腐蚀可与图像相减结合起来得到一幅图像的形态学梯度，形态学梯度由 g 定义：

$$g = (f \oplus b) - (f \ominus b) \tag{9.6-11}$$

膨胀粗化一幅图像中的区域，而腐蚀则细化它们。膨胀和腐蚀之差强调区域间的边界。同质区域不受影响(只要 SE 相对较小)，因此相减操作趋于消除同质区域。最终结果是是边缘被增强而同质区域的贡献则被抑制的图像，因此产生了“类似于微分”(梯度)的效果。

图 9.39 显示了一个例子。图 9.39(a)是一幅头部的 CT 扫描图像，接下来的两幅图像是使用所有元素都为 1 的 3×3 结构元对该图像进行开操作和闭操作的结果。注意刚刚提到的粗化和收缩。图 9.39(d)是使用式(9.6-11)得到的形态学梯度，其中清楚地描绘了区域间的边界，这与二维微分图像的预期相同。

a b c d

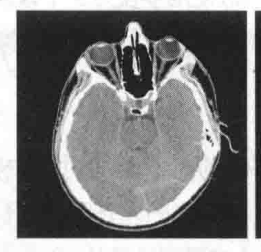
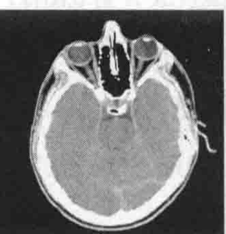
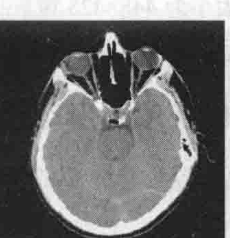
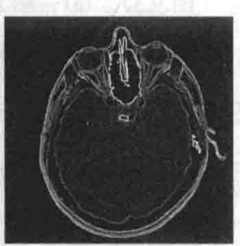





图 9.39 (a)512×512 的头部 CT 扫描图像；(b)膨胀的结果；(c)腐蚀的结果；(d)形态学梯度，计算图(b)和图(c)间的差(原图像由 Vanderbilt 大学的 David R. Pickens 博士提供)

692 ~ 694

顶帽变换和底帽变换

图像相减与开操作和闭操作相结合，可产生所谓的 Top-hat(顶帽)变换和 bottom-hat(底帽)变换。灰度级图像 f 的顶帽变换定义为 f 减去其开操作：

$$T_{\text{hat}}(f) = f - (f \circ b) \tag{9.6-12}$$

类似地，f 的底帽变换定义为 f 的闭操作减去 f：

$$B_{\text{hat}}(f) = (f \bullet b) - f \tag{9.6-13}$$

这些变换的主要应用之一是，用一个结构元通过开操作或闭操作从一幅图像中删除物体，而不是拟合被删除的物体。然后，差操作得到一幅仅保留已删除分量的图像。顶帽变换用于暗背景上的亮物体，而底帽变换则用于相反的情况。由于这一原因，当我们谈到这两个变换时，常常分别称为白顶帽变换和黑底帽变换。

顶帽变换的一个重要用途是校正不均匀光照的影响。如我们在下一章中看到的那样，合适(均匀)的光照在从背景中提取目标的处理中扮演核心的角色。称为分割的这一处理是自动图像分析中执行的第一步。一种常用的分割方法是对输入图像进行阈值处理。

为便于说明，考虑图9.40(a)，它显示了一幅大小为 600×600 的米粒图像。该图像是在非均匀光照下得到的，如图像底部及最右侧的暗色区域就是明证。图9.40(b)显示了对该图像使用10.3.3 节讨论的 Otsu 最佳阈值处理方法得到的结果。非均匀光照的最终结果导致了暗区域的分割错误(一些米粒未从背景中提取出来)，且在图像的左上角，背景部分被错误地分类了。图 9.40(c)显示了对该图像使用一个半径为 40 的圆盘形结构元进行开操作的结果。这个结构元大到足以不会拟合任何物体。如结果那样，这些物体被消除了，仅留下一个近似的背景。阴影模式在该图像中很清楚。通过从原图像中减去该图像(即执行顶帽变换)，背景应会变得更均匀。事实的确如此，如图9.40(d)所示。背景并不是非常均匀，但亮和暗之间不再存在极端的差别，这就足以得到正确的分割结果，其中所有的米粒均被检测出来，如图 9.40(e)所示。以前，该图像是使用 Otsu 方法得到的。

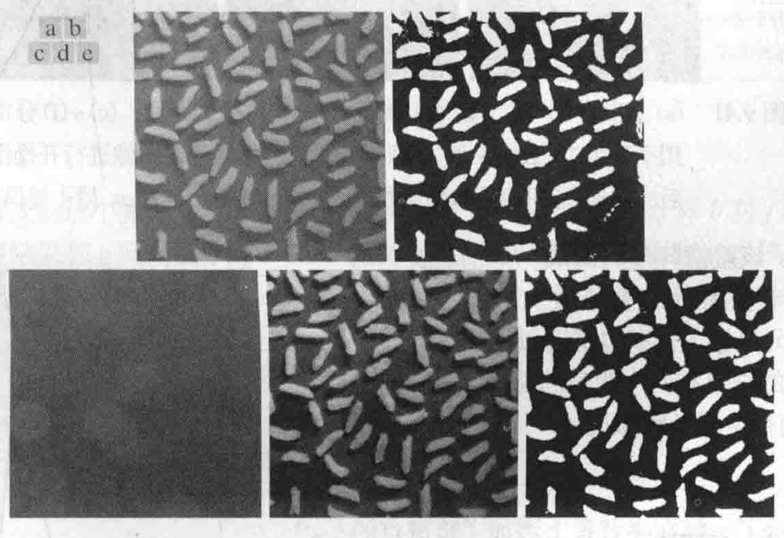


图 9.40　使用顶帽变换校正阴影：(a)大小为 600×600 的原图像；(b)阈值处理后的图像；(c)用半径为 40 的圆盘形结构元进行开操作的图像；(d)顶帽变换的图像(图像减去其开操作)；(e)顶帽变换的图像经阈值处理的结果

粒度测定

694 ~ 696

在图像处理方面，粒度测定属于判断图像中颗粒的尺寸分布的领域。实践中，颗粒很少能完

美地分离，这就使得用识别单个颗粒来计数很困难。形态学可间接用于估计颗粒的尺寸分布，而不需要识别并测量图像中的每个颗粒。

这种方法原理上非常简单。对于比背景亮且具有规则形状的颗粒，该方法由使用逐渐增大的结构元对图像执行开操作组成。基本概念是，某个特殊尺寸的开操作应会对包含类似尺寸的颗粒的输入图像的区域产生最大的效果。对于每次开操作，计算该开操作中像素值的和。该和有时称为表面区域，它会随着结构元的增大而减小，如我们早些时候注释的那样，因为开操作会降低亮特征的灰度。该过程会得到一个这样的数字的一维阵列，阵列中的每个元素等于对应于阵列中该位置的结构元素的大小的开操作中的像素之和。为了强调连续开操作间的变化，我们计算一维阵列的相邻元素的差。为了形象化该结果，我们画出了该差的图形。曲线中的峰值表明了图像中颗粒的主要大小分布。

作为一个例子，考虑图9.41(a)，它是两种不同大小的木钉图像。对木钉中的木颗粒可能需要引入不同的开操作，因此，平滑是一种合理的预处理步骤。图9.41(b)显示了使用早期讨论过的形态学滤波器平滑后的图像，该滤波器用半径为5的圆盘形结构元。图9.41(c)到(f)显示了使用半径为10, 20, 25和30的圆盘形结构元对图像进行开操作的例子。注意，在图9.41(d)中，由于较小的木钉，灰度的贡献几乎已被消除。图9.41(e)中，大木钉的贡献已被显著地降低，图9.41(f)更厉害［观察图9.41(e)，靠近图像右上方的大木钉比其暗得多，因为它的尺寸较小。如果我们试图检测有缺陷的木钉，那么这将是有用的信息］。

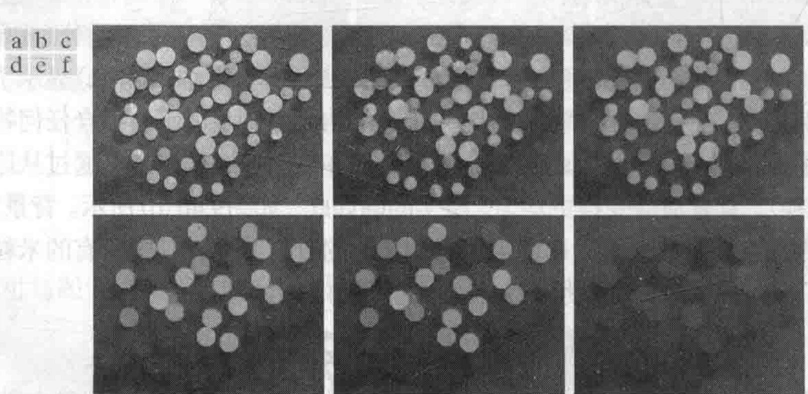


图9.41 (a)大小为531×675的木钉图像；(b)平滑后的图像；(c)~(f)分别用半径为10, 20, 25和30像素的圆盘形结构元对图像进行开操作后的结果(原图像由 MathWorks 公司的 Steve Eddins 博士提供)

图9.42显示了该差值阵列的曲线。如前边提到的那样，我们期望半径附近的差值较大(曲线中的峰值)，在该处结构元足够大，以包围近似相同直径的一组颗粒。图9.42中的结果有两个明显的峰值，这清楚地表明图像中存在两种主要的物体尺寸。

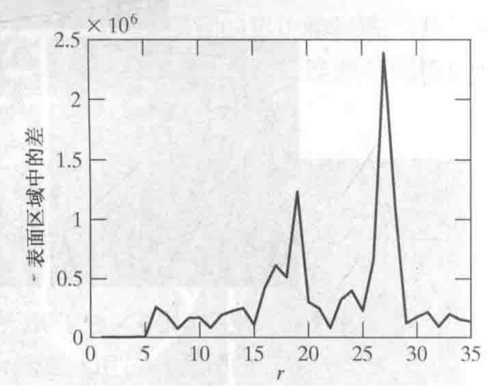


图9.42 表面区域中的差，它是圆盘形结构元半径 r 的函数。两个峰值表明图像中存在两个主要的颗粒尺寸

纹理分割

图9.43(a)显示了一幅在亮背景上添加了暗斑点的噪声图像。该图像有两个纹理区域：右边大斑点组成的区域和左边小斑点组成的区域。目的是以纹理内容为基础找到两个区域的边界(我们将在11.3.3节讨论纹理)。如早期解释的那样，将一幅图像分为区域的处理称为分割，分割是第10章的主题。

696 ~ 697

感兴趣的物体比背景暗，并且我们知道，如果使用一个比小斑点大的结构元对图像做闭操作，这些斑点将被删除。在图 9.43(b)中，使用半径为 30 像素的圆盘形结构元对输入图像做闭操作的结果，表明情况的确如此(斑点的半径近似为 25 像素)。因此，我们得到了一幅亮背景上带有暗色大斑点的图像。如果使用尺寸大于这些斑点间的间隔的结构元对该图像做开操作，则最终结果应该是这样一幅图像，在该图像中，斑点间的亮间隔已被删除，留下了暗色斑点，并且这些斑点之间的暗间隔现在已变得相同。图 9.43(c)显示了使用半径为 60 的圆盘形结构元得到的结果。

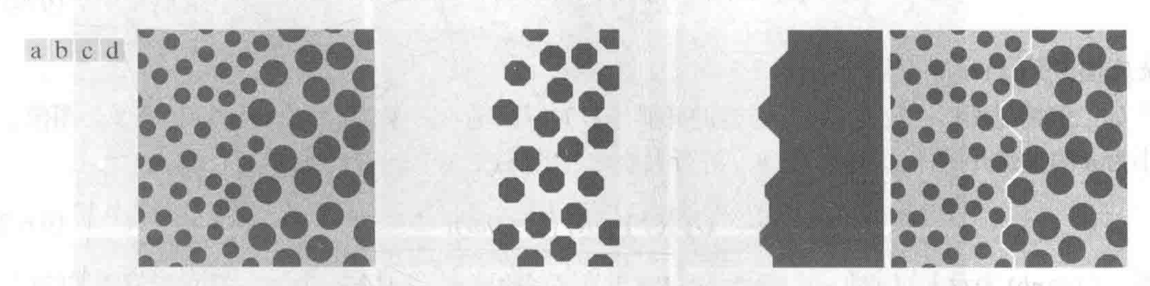


图 9.43　纹理分割：(a)由两种斑点组成的大小为 600×600 的图像；(b)对图(a)执行闭操作后删除了小斑点的图像；(c)对图(b)执行开操作后删除了大斑点间的亮间隔的图像；(d)将图(c)中两个区域间的边界叠加到原图像上后的结果。边界是使用形态学梯度操作得到的

使用一个元素全为 1 且大小为 3×3 的结构元对该图像执行形态学梯度操作，将给出这两个区域间的边界。图9.43(d)显示了将形态学梯度操作结果叠加到原图像后所得到的边界。该边界右侧的所有像素表示属于由大斑点表征的纹理区域，而该边界左侧的所有像素属于由小斑点表征的纹理区域。读者将发现，若使用图 9.36 中对开操作和闭操作的图形模拟，那么我们可更直观地理解这一例子。

9.6.4　灰度级形态学重建

灰度级形态学重建基本上按照与 9.5.9 节针对二值图像所介绍的相同方法来定义。令 f 和 g 分别代表标记图像和模板图像。我们假设 f 和 g 是大小相同的灰度级图像，且 $f \leqslant g$。f 关于 g 的大小为 1 的测地膨胀定义为

> 很容易理解这些表达式是 (x, y) 的函数。为简化表示，我们省略了 (x, y)。

$$D_g^{(1)}(f) = (f \oplus b) \wedge g \tag{9.6-14}$$

式中，$\wedge$ 代表点方式的最小算子。该式指出，大小为 1 的测地膨胀是先计算 b 对 f 的膨胀，然后选择在每个(x, y)点处该结果和 g 间的最小者。如果 b 是一个平坦结构元，则膨胀由式(9.6-2)给出，否则膨胀由式(9.6-4)给出。f 关于 g 的大小为 n 的测地膨胀定义为

$$D_g^{(n)}(f) = D_g^{(1)}\left[D_g^{(n-1)}(f)\right] \tag{9.6-15}$$

并有 $D_g^{(0)}(f) = f$。

类似地，f 关于 g 的大小为 1 的测地腐蚀定义为

$$E_g^{(1)}(f) = (f \ominus b) \vee g \tag{9.6-16}$$

式中，$\vee$ 表示点方式的最大算子。f 关于 g 的大小为 n 的测地腐蚀定义为

$$E_g^{(n)}(f) = E_g^{(1)}\left[E_g^{(n-1)}(f)\right] \tag{9.6-17}$$

并有 $E_g^{(0)}(f) = f$。

697 ~ 699

本节中表达式间的对偶关系列表见习题 9.33。

灰度级标记图像 f 对灰度级模板图像 g 的膨胀形态学重建，定义为 f 关于 g 的测地膨胀反复迭代，直至达到稳定；即

$$R_g^D(f) = D_g^{(k)}(f) \tag{9.6-18}$$

且 k 应使 $D_g^{(k)}(f) = D_g^{(k+1)}(f)$。$f$ 对 g 的腐蚀的形态学重建类似地定义为

$$R_g^E(f) = E_g^{(k)}(f) \tag{9.6-19}$$

且 k 应使 $E_g^{(k)}(f) = E_g^{(k+1)}(f)$。

如二值情况那样，灰度级图像重建的开操作首先腐蚀输入图像，并用它作为标记图像。图像 f 的大小为 n 的重建开操作定义为，先对 f 进行大小为 n 的腐蚀，再由 f 的膨胀重建；即

$$O_R^{(n)}(f) = R_f^D[(f \ominus nb)] \tag{9.6-20}$$

式中，$(f \ominus nb)$ 表示 b 对 f 的 n 次腐蚀，如 9.5.7 节解释的那样。回忆式(9.5-27)针对二值图像的讨论，重建开操作的目的是保护腐蚀后留下的图像分量的形状。

类似地，图像 f 的大小为 n 的重建闭操作定义为，先对 f 进行大小为 n 的膨胀，再由 f 的腐蚀重建；即

$$C_R^{(n)}(f) = R_f^E[(f \oplus nb)] \tag{9.6-21}$$

式中，$(f \oplus nb)$ 表示 b 对 f 的 n 次腐蚀。因为对偶性，图像的重建闭操作可以用图像的求补得到，先得到重建开操作，然后再求结果的补。最后，如下例所示，称为重建顶帽的一种有用技术是从一幅图像中减去其重建开操作。

例 9.11 使用形态学重建展平复杂的背景。

在这个例子中，我们用规格化图 9.44(a)中不规则图像背景的一些步骤来说明灰度级重建的用途，即仅留下恒定灰度背景上的正文。解决这一问题能很好地说明形态学概念的能力。我们从抑制各个按键顶部的水平反射开始。在图像中反射比任何单个字符都宽，因此，我们应该能在腐蚀操作中用一条长水平线通过执行重建的开操作抑制它们。该操作将产生包含按键及其反射的背景。从原图像减去这个背景(即执行一个重建顶帽操作)将从原图像中消除水平反射和变化的背景。

图 9.44(b)显示了在腐蚀操作中用一条大小为 1×71 像素的水平线对原图像执行重建开操作的结果。我们已经能够仅用一个开操作就除去字符，但得到的背景将不会如图 9.44(c)所示那样均匀(例如，比较两幅图像中各个按键之间的区域)。图 9.44(d)显示了从图 9.44(a)中减去图 9.44(b)的结果。如期望的那样，水平反射和背景的变化被抑制了。为便于比较，图 9.44(e)显示了执行刚才的顶帽变换的结果(如本节之前讨论的那样，即从图像中减去“标准的”开操作)。从图 9.44(c)中背景的特征可以预料，图 9.44(e)中的背景不像图 9.44(d)中的那样均匀。

下一步是从按键的边缘删除垂直反射，这在图 9.44(d)中非常明显。使用一个其宽度近似等于这些反射的线状结构元来执行一个重建开操作，我们可删除垂直反射(在这种情况下，结构元的宽度约为 11 像素)。图 9.44(f)显示了对图 9.44(d)执行该操作的结果。垂直反射被抑制，有用字符的垂直笔画也被细化(如 SIN 中的 I)，因此，我们不得不寻找一种恢复该字符的方法。由于被抑制的字符非常靠近其他字符，因此，如果我们水平地膨胀遗留下来的字符，则膨胀后的字符将与先前被抑制字符所占的区域重叠。使用大小为 1×21 的线状结构元膨胀图 9.44(f)后，得到的图 9.44(g)表明事实的确如此。

699 ~ 701

此时所有余下的问题是复原被抑制的字符。考虑图9.44(g)中膨胀后的图像和图9.44(d)中重建顶帽图像间点方式的最小值形成的一幅图像。图9.44(h)显示了最小值图像(虽然该结果看起来接近我们的目标，但 SIN 中仍遗漏了 I)。在灰度级重建中，通过将这幅图像作为标记图像，并将膨胀后的图像作为模板图像［见式(9.6-18)］，我们得到如图 9.44(i)所示的最终结果。这幅图像表明所有字符都已完全从原图像的不规则背景，包括按键的背景中提取出来。图 9.44(i)中的背景都是均匀的。

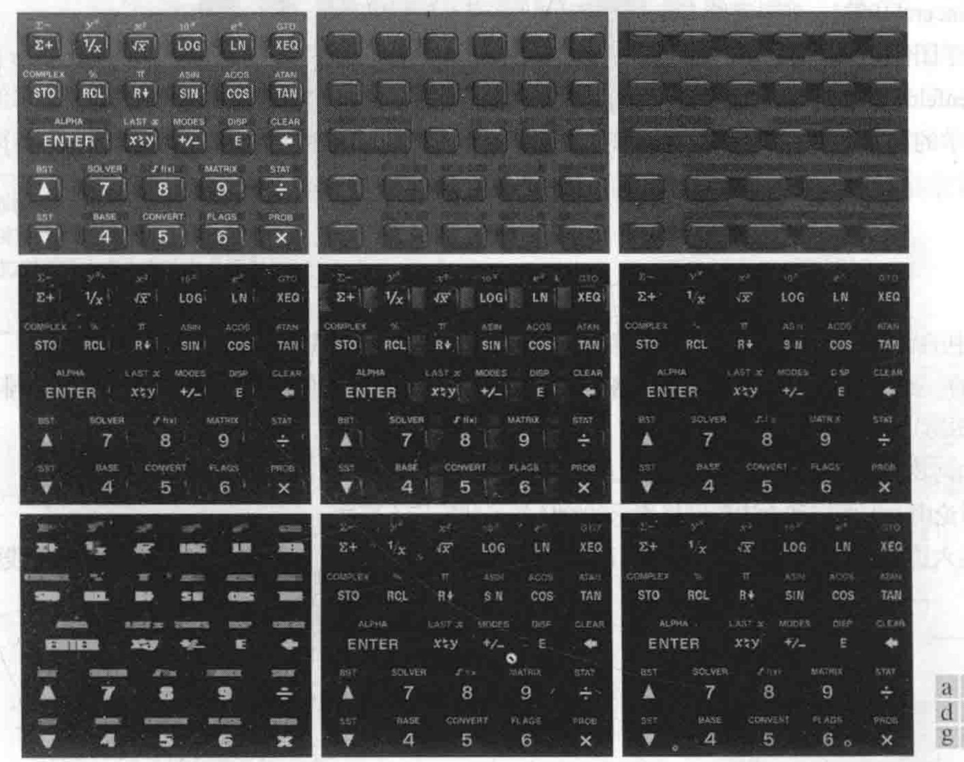


图 9.44　(a)大小为 1134×1360 像素的原图像；(b)用腐蚀中的一条 71 像素长的水平线对图(a)的重建开操作；(c)用相同的线对图(a)的开操作；(d)重建顶帽操作；(e)顶帽操作；(f)用 11 个像素长的水平线对图(d)的重建开操作；(g)使用 21 个像素长的水平线对图 f 的膨胀操作；(h)图(d)和图(g)的最小操作；(i)最后的重建结果(原图像由 MathWork 公司的 Steve Eddins 博士提供)

小结

本章介绍的形态学概念和技术构成了从一幅图像中提取感兴趣特征的一组有力工具。形态学图像处理最吸引人的一个方面是，来自形态学技术已有发展的集合论基础。实现方面的一个显著优点是，膨胀和腐蚀都是基本操作，它们是各类形态学算法的基础。如下章所示，形态学可以作为广泛应用的图像分割程序的基础。如第 11 章中讨论的那样，形态学技术在图像描述中也起着重要的作用。

参考文献

Serra[1982]是关于形态学图像处理的基础性参考资料，也可参阅 Serra[1988]、Giardina and Dougherty[1988]和Haralick and Shapiro[1992]。与讨论相关的其他早期参考文献包括 Blum[1967]、Lantuéjoul[1980]、Maragos[1987]

701

和 Haralick et al.[1987]。关于二值形态学和灰度级形态学的综述，见 Basart and gonzalez[1992]和 Basart et al.[1992]。这组参考文献为 9.1 节到 9.4 节涵盖的内容提供了丰富的背景材料。9.5 节和 9.6 节内容的较好综述见 Soille[2003]。

实现形态学算法（如 9.5 节和 9.6 节中给出的算法）的重要刊物见 Jones and Svalbe[1994]、Park and Chin [1995]、Sussner and Ritter[1997]、Anelli et al.[1998]和 Shaked and Bruckstein[1998]。实现灰度级形态学算法的实践细节见 Vincent[1993]，也可参阅 Gonzalez, Woods and Eddins[2004]。

关于形态学图像处理理论和应用的其他读物见 Goutsias and Bloomberg[2000]和 *Pattern Recongnition* [2000]，也可参阅 Rosenfeld[2000]。Marchand-Maillet and Sharaiha[2000]和 Ritter and Wilson[2001]是关于二值图像处理和图像代数学的重要文献。关于形态学技术在图像处理方面的应用的当前工作，请参阅 Kim[2005]和 Evans and Liu[2006]。

标有星号的习题的详细解答，可在本书的网站上找到。该网站中还包含有基于本章内容而建议的项目。

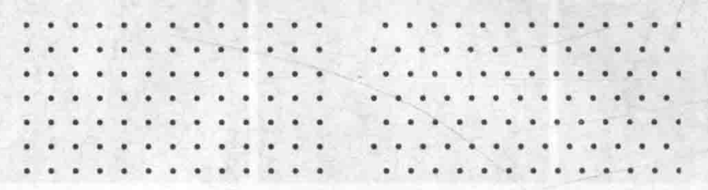
习题

9.1 本书中的数字图像嵌入了方形排列的网格，且像素允许是 4, 8 或 m 连通的。然而，也可使用其他的网格排列。特别地，有时也会使用 6 连通的六边形网格排列（见下面的图形）。

(a) 如何将图像从方形网格转换到六边形网格？

(b) 讨论由方形网格和六边形网格表示的物体的形状旋转不变性。

(c) 在六边形网格中，是否像在 8 连通情况下那样（见 2.5.2 节），可能存在有歧义的对角线配置？



9.2 ★(a) 针对把 4 连通的二值边界转换为 m 连通的边界（见 2.5.2 节）给出一个形态学算法。可以假设边界是全连通的，且边界的宽度为 1 像素。

(b) 你的算法的操作需要用每个结构元进行多于一次的迭代过程吗？ 试解释原因。

(c) 你的算法的性能是否与应用结构元的顺序无关？如果你的回答是无关，请证明之；否则请举出一个例子，说明你的过程与应用结构元的顺序有关。

9.3 只要 B 的原点包含于 B 中，结构元 B 对集合 A 的腐蚀就是 A 的一个子集。给出腐蚀 $A \ominus B$ 位于 A 之外或部分位于 A 之外的一个例子。

9.4 下列 4 种说法是正确的。请给出能证明这些说法正确的论据。(a)通常是正确的；(b)到(d)仅对数字集合才是正确的。为了证明(b)至(d)的正确性，请画出一个离散的方形网格（如习题 9.1 所示），并针对每种情况，使用由该网格上的点构成的集合给出一个例子。（提示：在仍能确定这些说法的正确性时，尽可能保持每种情况下的点数最少。）

★(a) 一个凸结构元对一个凸集的腐蚀仍是一个凸集。

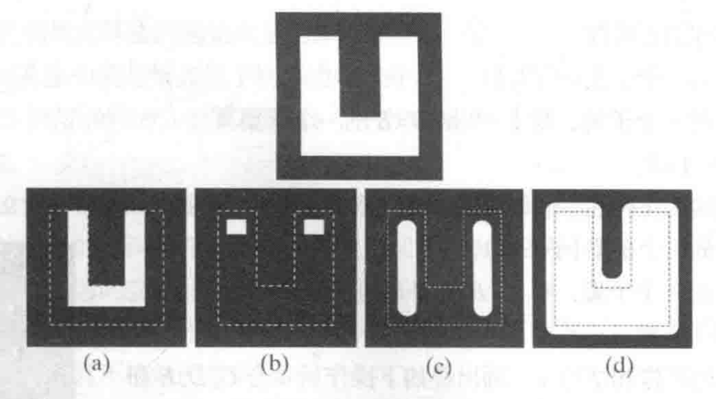
★(b) 一个凸结构元对一个凸集的膨胀并不总是凸的。

(c) 一个数字凸集内的点并不总是连通的。

(d) 有可能存在一个这样的点集，在该集合中，连接每个点对的线位于该集合内，但该集合不是凸的。

702~703

★9.5 参考所显示的图像，给出生成图(a)至图(d)所示结果的结构元和形态学操作。请清晰地说明每个结构元的原点。图中的虚线表示原始集合的边界，仅供参考显示。注意图(d)中的所有角都是圆角。



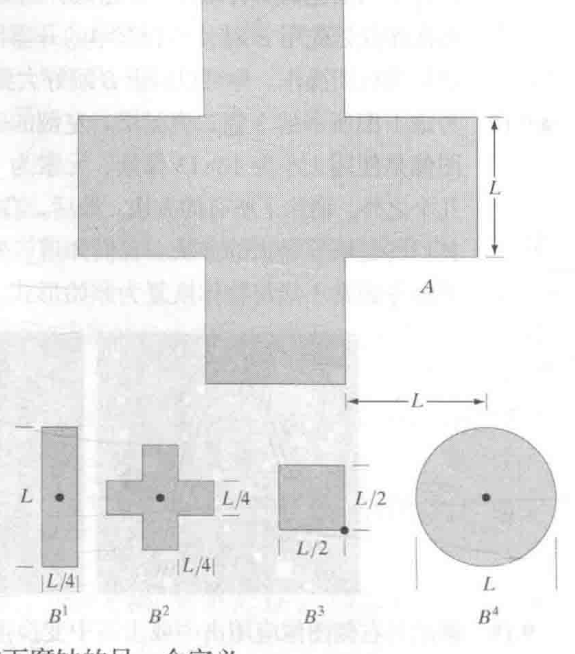
9.6 令 A 表示右图中阴影所示的集合。参考所示的结构元(黑点表示原点)，画出下列形态学操作的结果：

(a) $(A \ominus B^2) \oplus B^4$

(b) $(A \ominus B^1) \oplus B^3$

(c) $(A \ominus B^3) \oplus B^1$

(d) $(A \oplus B^3) \ominus B^2$



★9.7 (a) 反复膨胀一幅图像的极限效果是什么？假设不使用只有一个点的结构元。

(b) 为使(a)的答案成立，你能从其开始的最小图像是什么？

9.8 (a) 反复腐蚀一幅图像的极限效果是什么？假设不使用只有一个点的结构元。

(b) 为使(a)的答案成立，你能从其开始的最小图像是什么？

★9.9 腐蚀的另一种可替换定义是

$A \ominus B = \{w \in Z^2 \mid w + b \in A,\ 对于每个 b \in B\}$

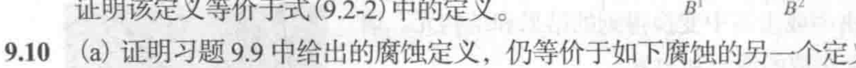
证明该定义等价于式(9.2-2)中的定义。

9.10 (a) 证明习题 9.9 中给出的腐蚀定义，仍等价于如下腐蚀的另一个定义：

$$A \ominus B = \bigcap_{b \in B} (A)_{-b}$$

(若用 b 代替 $-b$，则该表达式称为两个集合的闵可夫斯基相减。)

(b) 证明(a)中的表达式也等价于式(9.2-2)的定义。



★9.11 膨胀的一个可替换定义是

$$A \oplus B = \{w \in Z^2 \mid w = a + b,\ 对于一些\ a \in A\ 和\ b \in B\}$$

证明该定义等价于式(9.2-4)中的定义。

9.12 (a) 证明习题 9.11 中给出的膨胀定义，等价于如下关于膨胀的另一个定义：

$$A \oplus B = \bigcup_{b \in B} (A)_b$$

(该表达式也称为两个集合的闵可夫斯基相加。)

(b) 证明(a)中的表达式也等价于式(9.2-4)中的定义。

9.13 证明式(9.2-6)和式(9.2-5)中对偶表达式的正确性。

★9.14 证明对偶表达式 $(A \bullet B)^c = (A^c \circ \hat{B})$ 和 $(A \circ B)^c = (A^c \bullet \hat{B})$ 的正确性。

703 ~ 705

9.15 证明下列表达式的正确性：

★(a) $A \circ B$ 是 A 的一个子集(子图像)。

(b) 若 C 是 D 的一个子集，则 $C \circ B$ 是 $D \circ B$ 的一个子集。

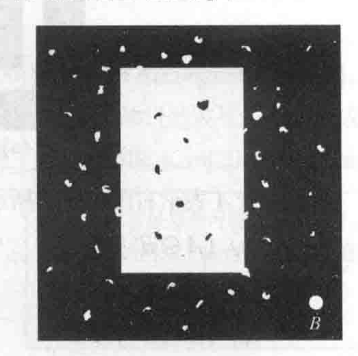
(c) $(A \circ B) \circ B = A \circ B$。

9.16 证明下列表达式的正确性［假设 B 的原点包含在 B 中，且习题 9.14 和习题 9.15 为真］：

(a) A 是 $A \bullet B$ 的一个子集(子图像)。

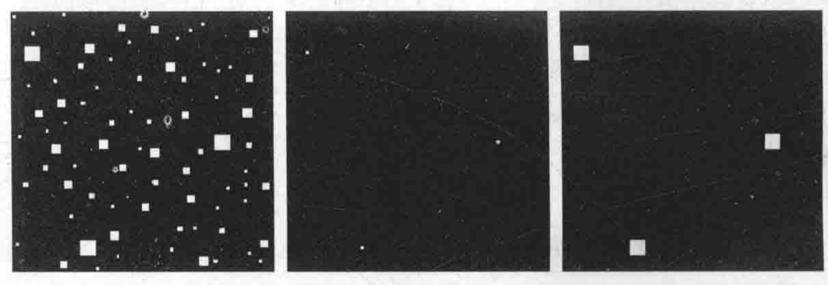
(b) 若 C 是 D 的一个子集，则 $C \bullet B$ 是 $D \cdot B$ 的一个子集。

(c) $(A \bullet B) \bullet B = A \bullet B$。



9.17 参考下图所示的图像和结构元。画出经如下操作后集合 C, D, E 和 F 是什么：$C = A \ominus B; D = C \oplus B; E = D \ominus B; F = E \oplus B$。初始集合 A 由显示为白色的所有图像分量组成，但结构元 B 除外。注意，这些操作仅是先用 B 对 A 进行简单的开操作，接着用 B 对开操作的结果进行闭操作。你可以假设 B 刚好大到包围每个噪声分量。

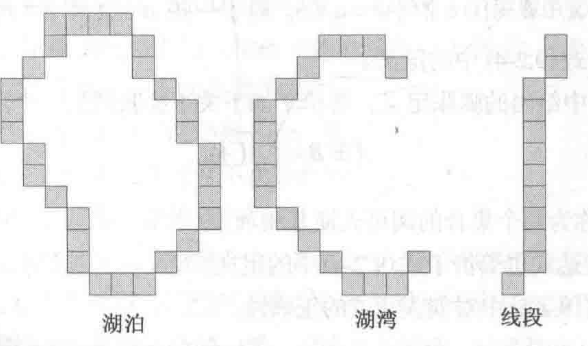
★9.18 考虑下图所示的 3 幅二值图像。左侧的图像由边长为 1, 3, 5, 7, 9 和 15 像素的方块组成。中间的图像是使用大小为 13×13 像素、元素为 1 的方形结构元对左侧图像进行腐蚀生成的，除了最大的几个之外，消除了所有的方块。最后，右侧的图像是使用相同的结构元对中间图像膨胀后的结果，其目的是恢复最大的方块。我们知道，先腐蚀再膨胀实际上是对图像的开操作，并且我们还知道开操作通常不能将物体恢复为原始形式。请解释这种情形下为何能完全重建较大的方块。





9.19 画出对右侧图像应用击中或击不中变换得到的结果和结构元。请清楚地指出你为结构元选取的原点和边界。

★9.20 用于区分图像中细化后物体的三个特征(湖泊、湖湾和线段)如下图所示。开发一种区分这些形状的形态学/逻辑算法。算法的输入是这三种形状之一。输出必须是输入的特征。可以假设这些特征的宽度都为 1 像素，且每个特征都是完全连通的，但它们出现时的方向可以是任意的。



705 ~ 707

9.21 在下列每种情况下，讨论你期望得到的结果是什么：

(a) 9.5.2 节中孔洞填充算法的起始点是物体边界上的一点。

(b) 该孔洞填充算法中的起始点位于物体边界之外。

(c) 使用 9.5.4 节中给出的算法，计算习题 9.6 中图形的凸壳外表特征是什么。请画出计算出的凸壳。假设 L 为 4 像素。

9.22★(a) 若使用图 9.15(c) 中给出的结构元来替代图 9.13(b) 中给出的结构元，试讨论对提取边界的影响。

(b) 在式(9.5-2)给出的孔洞填充算法中，若使用元素值为 1、大小为 2×2 的结构元代替图 9.15(c) 所示的结构元，请问会有什么影响？

9.23★(a) 针对图 9.16 中的自动化例子，请提出一种方法(使用 9.1 节到 9.5 节的任何技术)。可以假设球体互不接触，并且不接触图像的边界。

(b) 重复(a)，但允许球体以任意方式接触，包括接触边界。

★**9.24** 9.5.3 节中给出的提取连通分量的算法，要求在每个连通分量中有一个点是已知的，以便提取所有的点。假如你已有一幅包含任意数量(未知)连通分量的二值图像。请提出一种完全自动化地提取所有连通分量的过程。假设属于连通分量的点标记为 1，背景点标记为 0。

9.25 给出一个基于膨胀重建的、能提取一幅二值图像中的所有孔洞的表达式。

9.26 参考 9.5.9 节中的孔洞填充算法：

(a) 如果 f 的所有边界点都是 1，解释会发生什么情况。

(b) 如果(a)中的结果是你所希望的结果，请解释原因。如果不是，请解释如何改进该算法，以便算法能像你所期望的那样工作。

★**9.27** 如果结构元是值为 1 的单个点，解释二值图像的腐蚀和膨胀将会发生什么情况。对于你的回答，请给出原因。

9.28 如式(9.5-27)和 9.6.4 节中解释的那样，重建开操作可保持腐蚀后保留的图像分量的形状。请问重建闭操作的作用是什么？

★**9.29** 证明测地腐蚀和测地膨胀(见 9.5.9 节)关于补集对偶。也就是说，证明 $E_G^{(n)}(F)=[D_{G^c}^{(1)}[D_{G^c}^{(n-1)}(F^c)]]^c$，反之 $D_G^{(n)}(F)=[E_G^{(1)}[E_G^{(n-1)}(F^c)]]^c$。假定结构元关于其原点对称。

9.30 证明膨胀重建和腐蚀重建(见 9.5.9 节)关于补集对偶。也就是说，证明 $R_G^D(F)=[R_{G^c}^E(F^c)]^c$，反之亦然，$R_G^E(F)=[R_{G^c}^D(F^c)]^c$。假定结构元关于其原点对称。

★**9.31** 提出一个论据证明：

(a) $[(F\ominus nB)]^c=(F^c\oplus n\hat{B})$，其中 $(F\ominus nB)]$ 表示 B 对 F 的 n 次腐蚀。

(b) $[(F\oplus nB)]^c=(F^c\ominus n\hat{B})$。

9.32 证明二值重建闭操作与重建开操作关于补集对偶，即 $O_R^{(n)}(F)=[C_R^{(n)}(F^c)]^c$，并且类似地有 $C_R^{(n)}(F)=[O_R^{(n)}(F^c)]^c$。假设结构元关于其原点对称。

9.33 证明下列灰度级形态学表达式的正确性。你可以假设 b 是一个平坦结构元。回忆可知，$f^c(x,y)=-f(x,y)$ 且 $\hat{b}(x,y)=b(-x,-y)$。

★(a) 腐蚀和膨胀的对偶性：$(f\ominus b)^c=f^c\oplus\hat{b}$ 和 $(f\oplus b)^c=f^c\ominus\hat{b}$。

(b) $(f\bullet b)^c=f^c\circ\hat{b}$ 和 $(f\circ b)^c=f^c\bullet\hat{b}$。

★(c) $D_g^{(n)}(f)=[E_{g^c}^{(1)}[E_{g^c}^{(n-1)}(f^c)]]^c$ 和 $E_g^{(n)}(f)=[D_{g^c}^{(1)}[D_{g^c}^{(n-1)}(f^c)]]^c$。假设一个对称结构元。

(d) $R_g^D(f)=[R_{g^c}^E(f^c)]^c$ 和 $R_g^E(f)=[R_{g^c}^D(f^c)]^c$。

(e) $(f\ominus nb)^c=(f^c\oplus n\hat{b})$，其中 $(f\ominus nb)$ 表示 b 对 f 的 n 次腐蚀。此外，$[(f\oplus nb)]^c=(f^c\ominus n\hat{b})$。

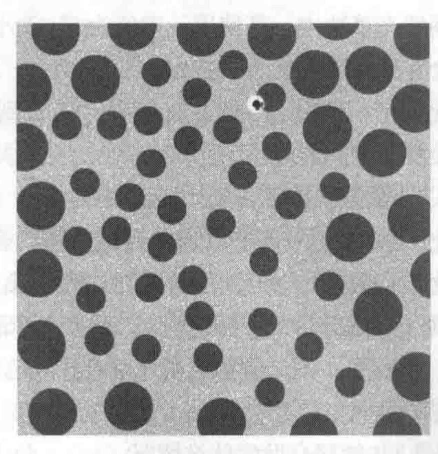
(f) $O_R^{(n)}(f)=[C_R^{(n)}(f^c)]^c$ 和 $C_R^{(n)}(f)=[O_R^{(n)}(f^c)]^c$。假设结构元关于其原点对称。

707 ~ 708

9.34 在图 9.43 中，建立两个不同纹理区域间的边界并不困难。考虑下图，它显示了一个由大圆区域包围的小圆区域。

(a) 用于生成图 9.43(d)的方法也可用于该图像吗？解释你的理由，包括为使且该方法进行处理而需要做出的任何假设。

(b) 如果你的回答是肯定的，请画出边界的外表特征。



9.35 一幅灰度级图像 $f(x, y)$被非重叠的噪声尖刺所污染，它们可以用半径为 $R_{\min} \leqslant r \leqslant R_{\max}$、幅度为 $A_{\min} \leqslant a \leqslant A_{\max}$ 的较小圆柱形干扰来建模。

★(a) 请开发一种清理该图像的形态学滤波方法。

(b) 重复(a)，但现在假设最多存在 4 个重叠的噪声尖刺。

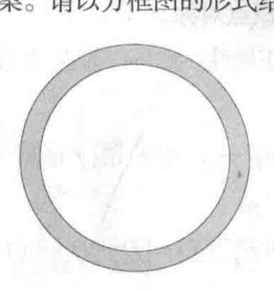
9.36 显微应用中的一个预处理步骤是，从两组或更多组重叠的类似颗粒(见右图)中分离出单个圆颗粒。假设所有颗粒的大小相同，提出一种产生 3 幅图像的形态学算法，这 3 幅图像分别仅由如下物体组成：

★(a) 仅与图像边界融合在一起的颗粒。

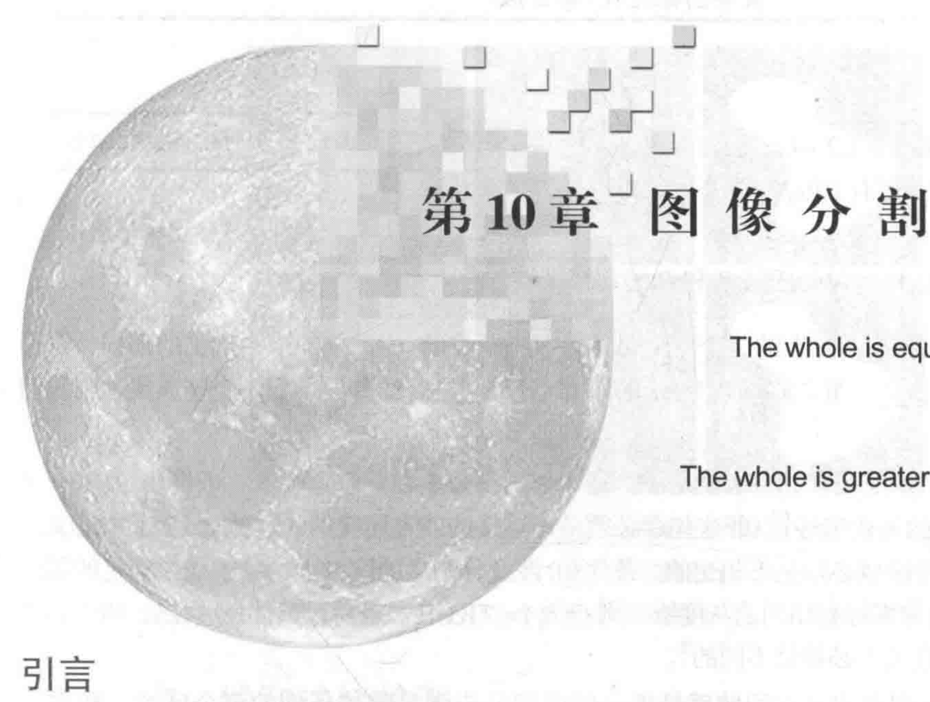
(b) 仅彼此重叠的颗粒。

(c) 不重叠的颗粒。

9.37 一家高技术制造工厂获得了一份制造形如下图的高精度垫圈的政府合同。合同要求使用图像系统来检测所有垫圈的形状。在合同文本中，形状检测是指检测垫圈内边缘和外边缘的偏差。你可以假定：(1) 一幅可接受的垫圈的“金”图像是可用的(关于该习题而言的一幅完美图像)；(2) 该系统中最终使用的成像和定位系统的精度，可高到足以允许你忽略数字化和定位引起的误差。假设你被聘为该系统视觉检测部分的顾问。请基于形态学/逻辑操作，提出一种检测方案。请以方框图的形式给出你的答案。



708~709



第10章 图像分割

The whole is equal to the sum of its parts.

Euclid

The whole is greater than the sum of its parts.

Max Wertheimer

引言

从前一章开始，所介绍的内容就从输入和输出都是图像的图像处理方法，转到了输入是图像而输出是从这些图像中提取出来的属性的图像处理方法(在1.1节中的定义意义上)。分割是该方向的另一个主要步骤。

分割将图像细分为构成它的子区域或物体。细分的程度取决于要解决的问题。也就是说，在应用中，当感兴趣的物体或区域已经被检测出来时，就停止分割。例如，在电子元件的自动检测中，我们关注的是分析产品的图像，客观地确定是否存在特定的异常现象，如元件缺失或连线断裂。超过识别这些元素所需细节的分割是没有意义的。

异常图像的分割是图像处理中最困难的任务之一。分割的精度决定了计算分析过程的最终成败。因此，应该对改进准确分割的可能性给予相当的关注。在某些情形下，比如在工业检测的应用中，至少对环境进行某些范围的控制还是可能的。有经验的图像处理系统设计师总是非常关注这种机会。在其他应用中，比如自主目标获取，系统设计者无法控制操作环境，常用方法是将注意力集中于传感器类型的选择上，在减少图像无关细节的影响的同时，增强感兴趣物体。一个很好的例子是军方使用红外线成像来检测热特性的物体，如运动中的装备和部队。

本章中的多数分割算法均基于灰度值的两个基本性质之一：不连续性和相似性。对于不连续的灰度，方法是以灰度突变为基础分割一幅图像，比如图像的边缘。对于相似的灰度，主要方法是根据一组预定义的准则把一幅图像分割为相似的区域。阈值处理、区域生长、区域分裂和区域聚合都是这类方法的例子。本章讨论和说明这些方法，并说明综合运用不同方法来改善分割的性能，例如，边缘检测就结合了阈值处理的技术。我们还将讨论基于形态学的图像分割。这种方法特别有吸引力，因为它基于本章第一节给出的技术，结合了分割的一些值得肯定的属性。最后，我们将以分割中运动线索运用的简单讨论结束本章。

> 基于多个灰度(亮度)值的分割技术的讨论，见6.7节和10.3.8节。

10.1 基础知识

令 R 表示一幅图像占据的整个空间区域。我们可以将图像分割视为把 R 分为 n 个子区域 $R_1, R_2, \cdots, R_n$ 的过程，满足：

(a) $\bigcup_{i=1}^{n} R_i = R$。

(b) R_i 是一个连通集，$i = 1, 2, \cdots, n$。

关于连通集的讨论，请参阅2.5.2节。

(c) $R_i \cap R_j = \varnothing$，对于所有 i 和 j，$i \neq j$。

(d) $Q(R_i) = \text{TRUE}$，$i = 1, 2, \cdots, n$。

(e) $Q(R_i \cup R_j) = \text{FALSE}$，对于任何 R_i 和 R_j 的邻接区域。

其中，$Q(R_k)$ 是定义在集合 R_k 的点上的一个逻辑属性，$\varnothing$ 表示空集。如2.6.4节中定义的那样，符号 $\cup$ 和 $\cap$ 分别表示集合的并和交。如2.5.2节中讨论的那样，若 R_i 和 R_j 的并集形成一个连通集，则我们说这两个区域是邻接的。

条件(a)指出，分割必须是完全的；也就是说，每个像素都必须在一个区域内。条件(b)要求一个区域中的点以某些预定义的方式来连接(即这些点必须是4连接的或8连接的，就像2.5.2节中定义的那样)。条件(c)指出，各个区域必须是不相交的。条件(d)涉及分割后的区域中的像素必须满足的属性——例如，若 R_i 中的所有像素都有相同的灰度级，则 $Q(R_i) = \text{TRUE}$。最后，条件(e)指出，两个邻接区域 R_i 和 R_j 在属性 Q 的意义上必须是不同的①。

这样，我们看到，分割中的基本问题就是把一幅图像分成满足前述条件的多个区域。通常，针对单色图像的分割算法处理两种类型的灰度值：不连续性和相似性。在第一类中，假设这些区域的边界彼此完全不同，且与背景不同，因此允许基于灰度的局部不连续性来进行边界检测。基于边缘的分割是这一类中所用的主要方法。第二类中基于区域的分割方法是根据事先定义的一组准则，把一幅图像分割成相似的几个区域。图10.1说明了前述概念。图10.1(a)显示了一幅在恒定灰度的暗背景上叠加一个恒定灰度区域的图像。这两个区域构成了整个图像区域。图10.1(b)显示了基于灰度的不连续性来计算内部区域的边界所得到的结果。边界内侧和外侧的点都是黑色(0)，因为在这些区域中灰度不存在不连续性。为了分割该图像，我们对该边界上或该边界内的像素分配了一个灰度级(譬如白)，而对该边界外部的所有点分配了另一个灰度(譬如黑)。图10.1(c)显示了这一处理方法的结果。我们看到，该结果满足本节开始时说明的条件(a)到(c)。条件(d)的属性是：若一个像素位于边界上或边界内，则将其标为白色；否则标为黑色。我们看到，在图10.1(c)中，对于标为黑色和白色的点，该属性为“真”。类似地，两个分割后的区域(物体和背景)满足条件(e)。

下面三幅图像说明了基于区域的分割。图10.1(d)类似于图10.1(a)，但内部区域的灰度形成了一幅纹理模式。图10.1(e)显示了计算该图像的边缘的结果。很明显，灰度中的大量寄生变化使得识别原图像中的唯一边界很困难，因为很多非零灰度变化连接到了边界上，故基于边缘的分割不是一种合适的方法。然而，我们注意到，外部区域是恒定的，因此，我们需要解决的这个简单分割问题的全部，就是区分纹理区域和恒定区域的一个属性。像素值的标准差是完成该任务的一种度量，因为在纹理区域标准差非零，而在其他区域标准差为零。图10.1(f)显示了将原图像分成大小为4×4的几个子区域后的结果。若某个子区域中的像素的标准差为正(即如果属性为“真”)，则将该子区域标记为白色；其他为零。结果是在区域边缘周围出现“块”效应，因为4×4方块组被标记为相同的灰度。最后，注意这些结果也满足本节开始处说明的5个条件。

712 ~ 714

① 一般来说，Q 可以是一个复合表达式，例如，如果 R_i 中像素的平均灰度小于 m_i，且如果它们的灰度的标准差大于 σ_i，其中 m_i 和 σ_i 是指定的常数，则 $Q(R_i) = \text{TRUE}$。

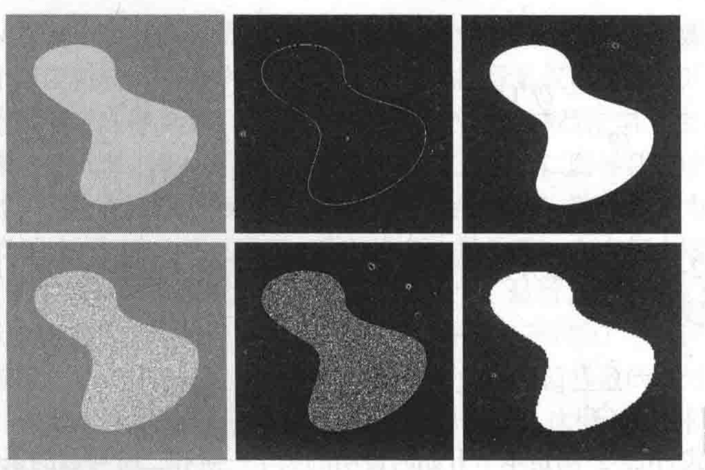


图10.1　(a)包含恒定灰度区域的图像；(b)显示内部区域边界的图像，该图像是由灰度不连续性获得的；(c)将图像分割成两个区域后的结果；(d)包含一个纹理区域的图像；(e)计算边缘后的结果。注意，由于存在大量连接到原始边界的小边缘，仅使用边缘信息是很难找到一条唯一的边界的；(f)基于区域特性的分割结果

10.2　点、线和边缘检测

本节主要介绍以灰度局部剧烈变化检测为基础的分割方法。我们感兴趣的三种图像特征是孤立点、线和边缘。边缘像素是图像中灰度突变的那些像素，而边缘(或边缘线段)是相连边缘像素的集合(关于连接性的定义，见2.5.2节)。边缘检测器是设计用来检测边缘像素的局部图像处理方法。一条线可视为一条边缘线段，该线两侧的背景灰度要么远亮于该线的像素的灰度，要么远暗于该线的像素的灰度。事实上，如下节及10.2.4节中讨论的那样，线会引起所谓的“屋顶边缘”。类似地，孤立点可视为一条线，只是其长度和宽度都为一个像素。

> 当我们说到线时，实际上是指那些较细的结构，它通常只有几个像素粗，譬如数字化后的建筑设计图中的线，或卫星图像中的道路。

10.2.1　背景知识

如在2.6.3节和3.5.1节我们所看到的局部平均平滑一幅图像那样，假设平均处理类似于积分，对于灰度的突变，局部变化可以用微分来检测应该并不奇怪。由于变化非常短促，因此一阶微分和二阶微分特别适合于这种目的。

数字函数的导数可用差分来定义。就像3.6.1节中解释的那样，近似这些差分的方法有多种，但对于用于一阶导数的任何近似，我们要求：(1)在恒定灰度区域必须为零；(2)在灰度台阶或斜坡开始处必须不为零；(3)在沿灰度斜坡点处也必须不为零。类似地，对于所用的二阶导数的近似，我们要求：(1)在恒定灰度区域必须为零；(2)在灰度台阶或斜坡的开始处和结束处必须不为零；(3)沿灰度斜坡必须为零。因为我们处理的是数字量，故它们的值是有限的，最大可能的灰度变化也是有限的，且在其上发生变化的最短距离是在两个邻接像素间的距离。

> 回忆2.4.2节可知，为表达清晰，图像样本间的增量定义为1，因此在式(10.2-1)的推导中使用了 $\Delta x = 1$。

我们按如下方式来得到一维函数$f(x)$在点x处的导数的近似：将函数$f(x+\Delta x)$展开为关于x的泰勒级数，令$\Delta x = 1$，且只保留该级数的线性项(见习题10.1)。结果是数字差分

$$\frac{\partial f}{\partial x} = f'(x) = f(x+1) - f(x) \tag{10.2-1}$$

当我们考虑两变量的图像函数$f(x, y)$时，为了表示的一致性，这里我们使用了偏微分。此时，我们将处理沿两个空间轴的偏微分。很明显，当函数f只有一个变量时，有$\partial f / \partial x = \mathrm{d}f / \mathrm{d}x$。

714~715

对式(10.2-1)关于 x 微分，我们得到一个二阶导数表达式：

$$\begin{aligned}\frac{\partial^2 f}{\partial x^2} &= \frac{\partial f'(x)}{\partial x} = f'(x+1) - f'(x) \\ &= f(x+2) - f(x+1) - f(x+1) + f(x) \\ &= f(x+2) - 2f(x+1) + f(x)\end{aligned}$$

式中，第二行来自式(10.2-1)。这一展开是关于点 $x+1$ 的。我们的兴趣是关于点 x 的二阶导数，故将上式中的变量减 1 后，得到

$$\frac{\partial^2 f}{\partial x^2} = f''(x) = f(x+1) + f(x-1) - 2f(x) \tag{10.2-2}$$

很容易证明式(10.2-1)和式(10.2-2)满足本节开始时说明的关于一阶和二阶导数的条件。为了说明这一点，并强调一阶导数和二阶导数在图像处理中的基本相同点和不同点，考虑图 10.2。

图 10.2(a)显示了一幅图像，该图像中包含有不同的实心物体、一条线和单个噪声点。图 10.2(b)显示了近似通过该图像中心的水平灰度剖面(扫描线)，其包括孤立点。实心物体和沿扫描线的背景之间的灰度过渡显示了两种类型的边缘：斜坡边缘(左侧)和台阶边缘(右侧)。如稍后我们要讨论的那样，灰度过渡涉及较细物体，如常称为*屋顶边缘*的线。图10.2(c)显示了剖面线的简化，构成曲线的点刚好足够我们在遇到一个噪声点、一条线和物体的边缘时，在数量上分析一阶导数和二阶导数的特性。在这幅简化图中，斜坡的过渡跨越 4 个像素，噪声点是单个像素，线为 3 个像素粗，灰度台阶的过渡发生在邻接像素间。为简化起见，灰度级数限制为 8 级。

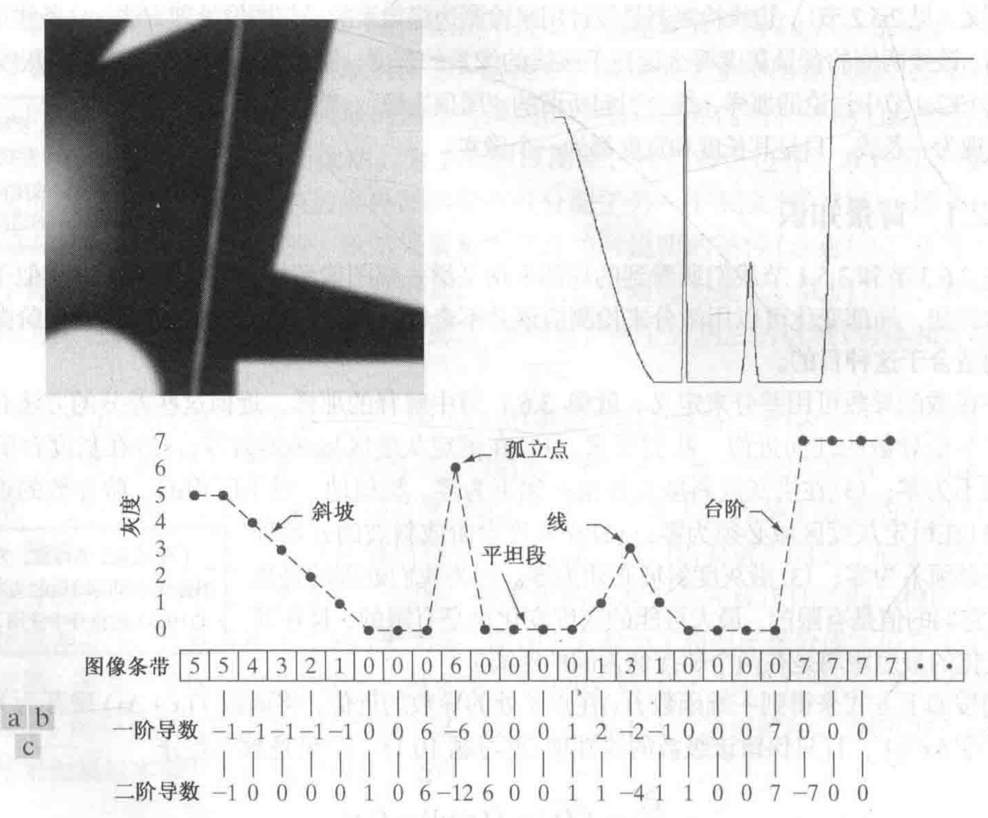


图 10.2 (a)图像；(b)通过该图像中心的水平灰度剖面，包括孤立噪声点；(c)简化后的剖面线(为清楚起见，点已用虚线连接起来)。图像条带对应于灰度剖面曲线，方框中的数字是剖面线中所示的点的灰度值。导数是使用式(10.2-1)和式(10.2-2)得到的

715 ~ 716

现在我们考虑从左到右横穿该剖面时一阶导数和二阶导数的性质。首先，我们注意到在灰度斜坡的开始处并沿着整个灰度斜坡，一阶导数不为零，同时二阶导数仅在斜坡的开始处和结尾处不为零。因为数字图像的边缘类似于这种类型的过渡，故我们断定一阶导数会产生“粗”边缘，而二阶导数则会产生细得多的边缘。接下来，我们遇到孤立噪声点。该点处的二阶导数响应幅度远强于一阶导数响应幅度。这并不意外，因为在增强剧烈变化方面，二阶导数要比一阶导数更为激进。这样，我们就可以预期，在增强细节(包括噪声)方面，二阶导数远强于一阶导数。在本例中线是如此之细，以至于它也是更精细的细节，并且我们再次看到二阶导数的幅度更大。最后，我们注意到在斜坡和台阶边缘中，进入边缘和离开边缘过渡时的二阶导数的符号相反(从负到正或从正到负)。如我们在 10.2.6 节中显示的那样，这种“双边缘效应”是可用于定位边缘的一种重要特性。二阶导数的符号也可用于确定一个边缘是从亮到暗(负二阶导数)过渡还是从暗到亮(正二阶导数)过渡，这里，当我们移进边缘时，就观察到这个符号。

总之，我们得出如下结论：(1)一阶导数通常在图像中产生较粗的边缘；(2)二阶导数对精细细节，如细线、孤立点和噪声有较强的响应；(3)二阶导数在灰度斜坡和灰度台阶过渡处会产生双边缘响应；(4)二阶导数的符号可用于确定边缘的过渡是从亮到暗还是从暗到亮。

计算图像中每个像素位置处的一阶导数和二阶导数的另一种方法是使用空间滤波器。对于图 10.3 中的 3×3 滤波器模板，该过程是计算模板系数与被该模板覆盖区域中的灰度值的乘积之和。也就是说，参考式(3.4-3)，模板在该区域中心点处的响应是

$$R = w_1 z_1 + w_2 z_2 + \cdots + w_9 z_9 = \sum_{k=1}^{9} w_k z_k \tag{10.2-3}$$

其中 z_k 是像素的灰度，该像素的空间位置对应于模板中第 k 个系数的位置。3.4 节和 3.6 节已详细讨论了对图像中所有像素执行这种操作的细节。换句话说，基于空间模板的导数的计算，是用这些模板对一幅图像进行空间滤波，正如在那些章节描述的那样[①]。

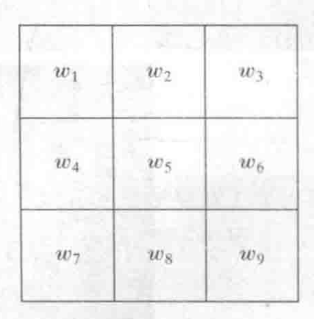


图 10.3　一个普通的 3×3 空间滤波器掩模

10.2.2　孤立点的检测

基于前一节得到的结论，我们可知点的检测应以二阶导数为基础。由 3.6.2 节的讨论，这意味着使用拉普拉斯算子：

$$\nabla^2 f(x, y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{10.2-4}$$

式中，偏微分用式(10.2-2)得到：

$$\frac{\partial^2 f(x, y)}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \tag{10.2-5}$$

和

$$\frac{\partial^2 f(x, y)}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \tag{10.2-6}$$

① 像 3.4.3 节中解释的那样，式(10.2-3)要么是由式(3.4-1)给出的空间相关的简化表示，要么是由式(3.4-2)给出的空间卷积的简化表示。因此，当在一幅图像中的所有位置处计算 R 时，结果都是一个阵列。本章中的所有空间滤波操作都是使用相关完成的。在某些情形下，我们通常会使用术语“模板与图像卷积”。但是，我们仅在滤波器是对称滤波器时才使用这一术语，因为这一情形下相关和卷积会产生相同的结果。

716 ~ 718

故拉普拉斯算子是

$$\nabla^2 f(x, y) = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y) \tag{10.2-7}$$

如 3.6.2 节中解释的那样，该表达式可以用图 3.37(a)中的模板来实现。此外，如该节描述的那样，我们可以把式(10.2-7)扩展到包含对角项，并使用图3.37(b)中的模板。使用图10.4(a)中的拉普拉斯模板，它与图3.37(b)中的模板相同，如果在某个点处，该模板的响应的绝对值超过了某个指定的阈值，那么我们说在模板中心位置(x, y)处的该点已被检测到。在输出图像中，这样的点被标注为 1，而所有其他点则被标注为 0，从而产生一幅二值图像。换句话说，输出是使用如下表达式得到的：

$$g(x, y) = \begin{cases} 1, & |R(x, y)| \geqslant T \\ 0, & \text{其他} \end{cases} \tag{10.2-8}$$

式中，g 是输出图像，T 是一个非负的阈值，R 由式(10.2-3)给出。该式简单地度量一个像素及其 8 个相邻像素间的加权差。从直观上看，这一概念是一个孤立点的灰度将完全不同于其周围像素的灰度，因而，使用这种类型的模板可很容易地检测出这个孤立点。考虑的重点仅仅是灰度的不同对于研究孤立点已很充分。注意，对于一个导数模板而言，这些系数之和为零通常表明在恒定灰度区域模板响应将是 0。

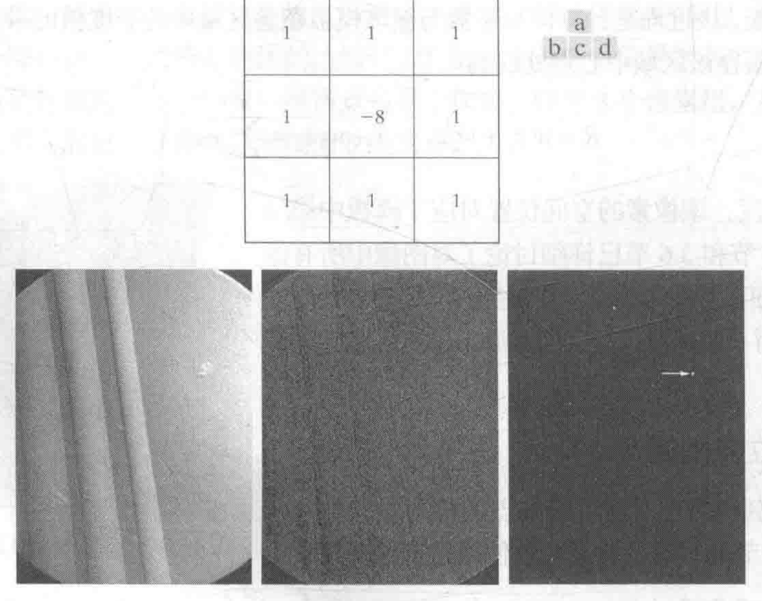


图 10.4 (a)点检测(拉普拉斯) 模板；(b)带有一个通孔的涡轮叶片的 X 射线图像。该通孔含有一个黑色像素；(c)模板与图像卷积的结果；(d)使用式(10.2-8)得到的结果，结果中显示了一个点(为便于观看，该点已被放大)(原图像由 X-TEK Systems 有限公司提供)

例 10.1 图像中孤立点的检测。

借助于图 10.4(b)，我们来说明如何从一幅图像中将孤立点分割出来。该图像是一幅喷气发动机涡轮叶片的 X 射线图像。图像右上象限的叶片中有一个通孔，通孔中嵌入了一个黑色像素。图 10.4(c)是将点检测模板应用到该幅 X 射线图像后的结果，图 10.4(d)显示了 T 取图 10.4(c)中像素的最高绝对值的 90%时，应用式(10.2-8)所得到的结果。在这幅图像中，这个像素清晰可见(为增加可视性，该像素已被人为放大)。这种类型的检测过程相当特殊，因为它基于单个像素位置处的灰度突变，这些位置被检测模板区域中的同质背景所围绕。当这一条件不满足时，本章中讨论的其他方法会更适合于检测灰度变化。

718~719

10.2.3 线检测

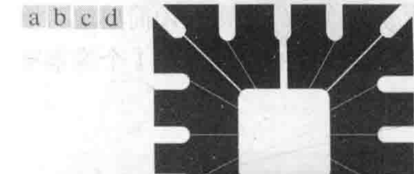
复杂度更高的检测是线检测。基于10.2.1节中的讨论，我们知道，对于线检测，可以预期二阶导数将导致更强的响应，并产生比一阶导数更细的线。这样，对于线检测，我们也可使用图10.4(a)中的拉普拉斯模板，记住，必须适当处理二阶导数的双线效应。下面的例子说明了这一处理过程。

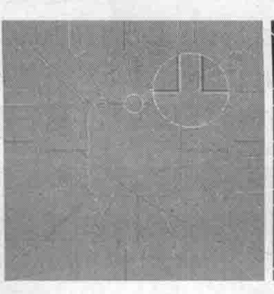
例10.2 用拉普拉斯进行线检测。

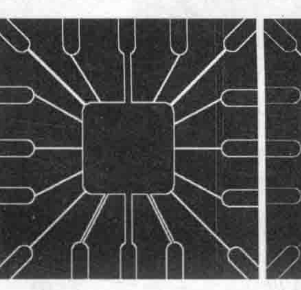
图10.5(a)显示了一幅486×486电路连线掩模的一部分(二值图像)，图10.5(b)显示了其拉普拉斯图像。因为拉普拉斯图像包含有负值[①]，为便于显示，需要调整比例。如放大部分显示的那样，中等灰度表示零，较暗的灰色调表示负值，而较亮的色调表示正值。在放大部分，双线效应清晰可见。

首先，负值看起来可通过取拉普拉斯图像的绝对值来简单地处理。然而，如图10.5(c)所示，该方法会使线的宽度加倍。一种更合适的方法是，仅使用拉普拉斯的正值(在有噪声的情形下，我们使用超过正阈值的那些值，去掉那些由噪声导致的零附近的随机变化)。如图10.5(d)中的图像所示，这种方法产生了更细的线，但这些线更有用。注意，在图10.5(b)到图10.5(d)中，当线的宽度比拉普拉斯模板的尺寸宽时，这些线就被一个零值“山谷”分开。

a b c d







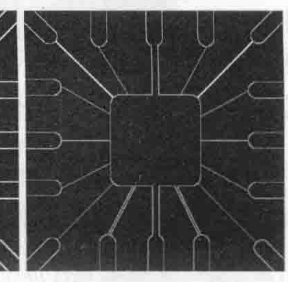


图10.5 (a)原图像；(b)拉普拉斯图像，放大部分显示了拉普拉斯的正/负双线效应；(c)拉普拉斯的绝对值；(d)拉普拉斯的正值

这是预料之中的。例如，当把这个3×3滤波器居中放在一条宽为5个像素的恒定灰度线上时，其响应将为零，这样就产生了刚才提到的效应。当我们谈论线检测时，假设这些线要细于检测子的尺寸。不满足这一假设的线，我们最好把它当作区域，并使用本节稍后讨论的边缘检测方法来处理。

图10.4(a)中的拉普拉斯检测子是各向同性的，因此其响应与方向无关(相对于该3×3拉普拉斯模板的4个方向：垂直方向、水平方向和两个对角方向)。通常，我们的兴趣在于检测特定方向的线。考虑图10.6中的模板。假设使用第一个模板对一幅具有恒定背景并包含各种线(方向为0°，±45°和90°)的图像进行滤波。最大响应将出现在图像中通过该模板中间行的一条水平线处。通过勾画一个元素为1的简单阵列，并画一条水平穿越该阵列的具有不同灰度(假设为5)的线，可很容易地验证这一点。类似的实验表明，图10.6中的第二个模板对于45°方向的线有最佳响应；第三个模板对于垂直线有最佳响应；第四个模板对于−45°方向的线有最佳响应。每个模板的首选方向用一个比其他方向更大的系数(如2)加权。每个模板中的系数之和为零，这表明恒定灰度区域中的响应为零。

回忆2.4.2节可知图像轴约定如下：原点位于左上角，正x轴指向下方，正y轴指向右方。本节中讨论的线的角度是指相对于正x轴度量的角度。例如，垂直线的角度为0°，+45°线向右下方向延伸。

① 当其系数之和为零的掩模与一幅图像进行卷积时，结果图像中的像素之和也为零(见习题3.16)，这表明结果中存在正像素和负像素。比例调节的目的是，使所有的负值适合于显示。

719 ~ 721

-1	-1	-1
2	2	2
-1	-1	-1

水平

2	-1	-1
-1	2	-1
-1	-1	2

+45°

-1	2	-1
-1	2	-1
-1	2	-1

垂直

-1	-1	2
-1	2	-1
2	-1	-1

−45°

图 10.6　线检测模板。角度是相对于图 2.18(b)中的坐标轴系统的

令 R_1, R_2, R_3 和 R_4 表示图 10.6 中从左到右的各个模板的响应，其中 R 值由式(10.2-3)给出。假设使用这 4 个模板对一幅图像滤波。在该图像中的某个给定点处，若对于所有 $j \neq k$ 有$|R_k| > |R_j|$，则称该点可能与模板 k 方向的一条线更相似。例如，若在图像中的某个点处，对于 $j = 2, 3, 4$ 有$|R_1| > |R_j|$，则说该点可能与一条水平线更相似。换句话说，我们可能对检测特定方向上的线感兴趣。在这种情况下，我们会使用与该方向相关的模板，并对其输出进行阈值处理，就像在式(10.1-2)中那样。换句话说，若我们对检测图像中由给定模板定义的方向上的所有线感兴趣，则只须简单地对该图像运行这个模板，并对结果的绝对值进行阈值处理。留下的点是有最强响应的点，对于 1 个像素宽度的线来说，相应的点最接近于模板定义的方向。下例说明了这一过程。

注意，这里使用 R 来表示模板响应，而 10.1 节中则用来表示区域，请不要混淆。

例 10.3　特定方向线的检测

图 10.7(a)显示了前一例子中所用的图像。假设我们的兴趣在于寻找所有宽度为 1 个像素、方向为 45°的线。为了这一目的，我们使用图 10.6 中的第二个模板。图 10.7(b)是用该模板对图像滤波后的结果。如之前那样，图 10.7(b)中比背景暗的色调对应于负值。图像中有两个取向为+45°方向的主要线段：一个在左上方，另一个在右下方。图 10.7(c)和(d)显示了图 10.7(b)中对应于这两个区域的放大部分。注意，图 10.7(d)中的直线段比图 10.7(c)中的该线段亮得多。原因是图 10.7(a)中右下方线段的宽度为 1 像素，而左上方线段的宽度则不是。该模板被“调谐”到检测+45°方向的 1 像素宽的线，我们期望在检测这样的线时，其响应较强。图 10.7(e)显示了图 10.7(b)的正值。因为我们的兴趣在于最强响应，我们令 T 等于图 10.7(e)中的最大值。图 10.7(f)使用白色显示了其值满足条件 $g \geqslant T$ 的点，其中 g 是图 10.7(e)中的图像。该图中的孤立点是对模板也有类似强响应的点。在原图像中，这些点与它们的邻点都按这样一种方法来取向，即模板在这些位置会产生最大的响应。用图10.4(a)中的模板可检测这些孤立点，然后删除这些点；或者可以使用上一章中讨论的形态学算子来删除这些孤立点。

10.2.4　边缘模型

边缘检测是基于灰度突变来分割图像的常用方法。我们先介绍一些边缘建模的方法，然后讨论一些边缘检测方法。

边缘模型根据它们的灰度剖面来分类。台阶边缘是指在 1 像素的距离上出现两个灰度级间的理想过渡。图 10.8(a)显示了一个垂直台阶边缘的一部分和通过该边缘的一个水平剖面。例如，用于诸如固体建模和动画领域的、出现在由计算机生成的图像中的台阶边缘。这些清晰、理想的边缘可出现在 1 像素的距离上，不需要提供任何使它们看上去“很真实”的附加处理(如平滑)。在算法开发中，数字台阶边缘常用做边缘模型。例如，10.2.6 节中讨论的坎尼边缘检测算法就是用一个台阶边缘模型推导的。

721 ~ 722

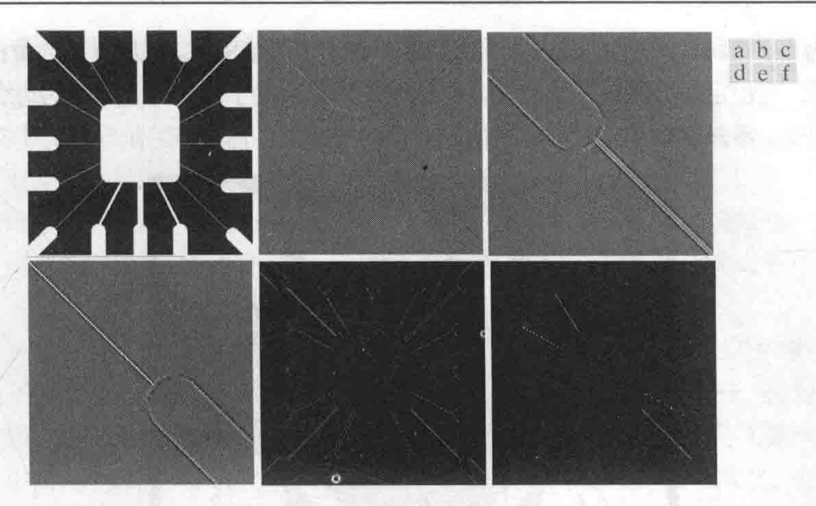


图 10.7　(a)接线板掩模图像；(b)使用图 10.6 中的+45°线检测子模板处理后的结果；(c)图(b)左上方区域的放大观察图；(d)图(b)右下区域的放大观察图；(e)将图(b)中所有负值置为零后的图像；(f)其值满足条件 $g \geqslant T$ 的所有点(白色)，其中 g 是图(e)中的图像［为便于查看，图(f)中的点已被放大］

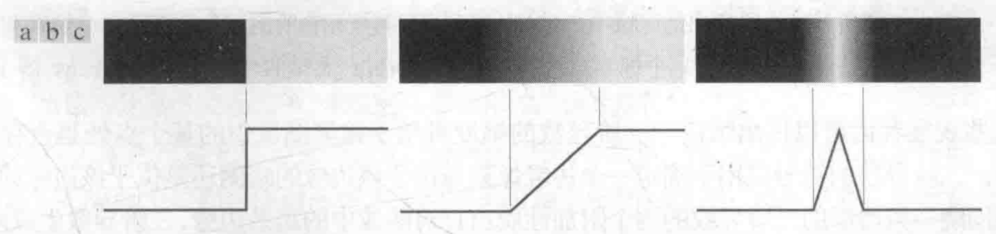


图 10.8　从左到右分别为一个台阶模型、一个斜坡模型和一个屋顶边缘模型(理想表示)及它们的相应灰度剖面

实际中，数字图像都存在被模糊且带有噪声的边缘，模糊的程度主要取决于聚焦机理(如光学成像中的镜头)中的限制，而噪声水平主要取决于成像系统的电子元件。在这种情况下，边缘被建模为一个更接近灰度斜坡的剖面，如图10.8(b)中的边缘。斜坡的斜度与边缘的模糊程度成反比。在这一模型中，不再存在一条细的(1 像素宽) 轨迹。相反，一个边缘点现在是斜坡中包含的任何点，而一条边缘线段将是一组已连接起来的这样的点。

边缘的第三种模型是所谓的“屋顶”边缘，这种边缘具有图10.8(c)所示的特性。屋顶边缘是通过一个区域的线的模型，屋顶边缘的基底(宽度)由该线的宽度和尖锐度决定。在极限情形下，当其基底为 1 像素宽时，屋顶边缘只不过是一条穿过图像中一个区域的一条 1 像素宽的线。例如，在深度成像中，当细物体(如管子)比等距离的背景(如墙)更接近传感器时，就会出现屋顶边缘。管道更亮，因而产生了一幅类似于图10.8(c)中模型的图像。如先前提及的那样，经常出现屋顶边缘的其他领域是在数字化的线条图和卫星图像中，此时如道路这样的较细特征可由这种类型的边缘建模。

包含所有三种类型边缘的图像并不罕见。虽然模糊和噪声会导致与理想形状的偏差，但图像中有适当锐度和适中噪声的边缘确实存在类似于图10.8 中边缘模型的特性，如图10.9 所示的剖面[①]。图10.8 中的模型允许我们在图像处理算法的开发中写出边缘的数学表达式。这些算法的性能将取决于实际边缘和在算法开发中所用模型之间的差别。

图10.10(a)显示了从图 10.8(b)的一段中提取出来的一幅图像。图10.10(b)显示了一条水平灰度剖面线。该图形还显示了灰度剖面的一阶导数和二阶导数。如 10.2.1 节中讨论的那样，当沿着灰度剖面从左到右移动时，我们注意到，在斜坡的开始处和在斜坡上的各个点处，一阶导数为正。而在恒定灰度

722
~
725

① 有着几个像素的剧烈斜率的斜坡边缘，通常被当做台阶边缘处理，目的是由同一图像中斜率更为舒缓的斜坡来对它们进行微分。

区域的一阶导数为零。在斜坡的开始处，二阶导数为正；在斜坡的结束处，二阶导数为负；在斜坡上的各点处，二阶导数为零；在恒定灰度区域的各点处，二阶导数为零。对于从亮到暗过渡的边缘，刚刚讨论的导数的符号则正好相反。零灰度轴和二阶导数极值间的连线的交点，称为该二阶导数的零交叉点。

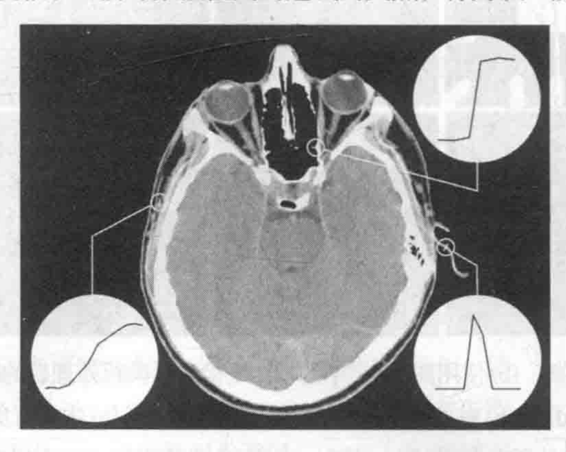


图 10.9　显示了(放大后的)实际斜坡(左下)、台阶(右上)和屋顶边缘剖面的一幅大小为 1508×1970 的图像。在由小圆所示短线段指出的区域中，剖面从暗到亮。斜坡和台阶剖面分别跨越 9 个像素和 2 个像素。屋顶边缘的基底是 3 个像素(原图像由 Vanderbilt 大学的 David R. Pickens 博士提供)

由这些观察我们可以得出结论：一阶导数的幅度可用于检测图像中的某个点处是否存在一个边缘。同样，二阶导数的符号可用于确定一个边缘像素是位于该边缘的暗侧还是位于该边缘的亮侧。我们注意到围绕一条边缘的二阶导数的两个附加性质：(1)对图像中的每条边缘，二阶导数生成两个值(一个不希望的特点)；(2)二阶导数的零交叉点可用于定位粗边缘的中心，正如我们将在本节稍后将要说明的那样。某些边缘模型可用来在进入和离开斜坡的地方平滑过渡(见习题 10.7)。然而，使用这些模型得到的结论与使用一个理想斜坡得到的结论相同，且用后者可简化理论公式。最后，尽管迄今为止我们的注意力限制在一维水平剖面上，但类似的结论适用于图像中任何方向的边缘。我们简单地定义了一个在任何期望的点与边缘方向垂直的剖面，并用刚讨论的与垂直边缘相同的方法，对结果进行了解释。

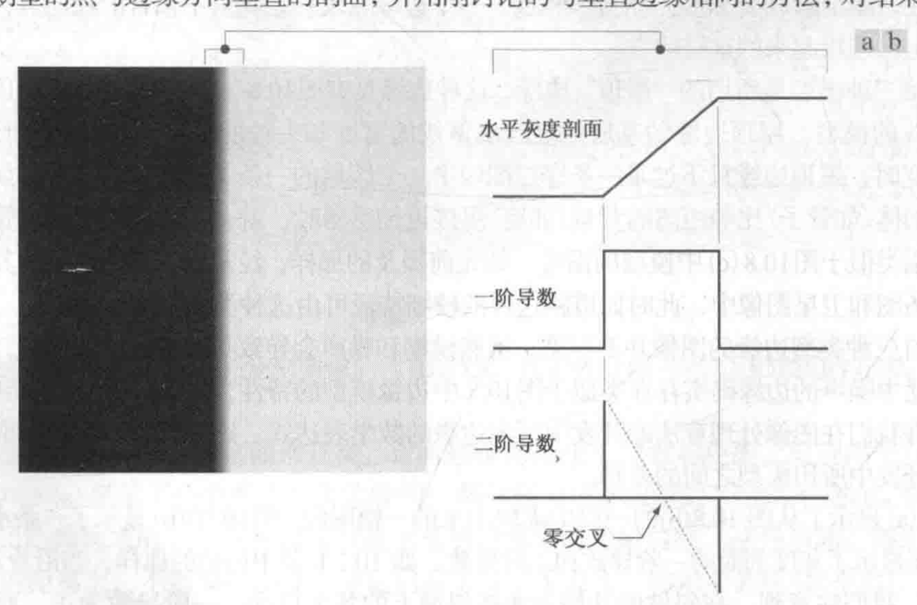


图 10.10　(a)由一条理想垂直边缘分开的两个恒定灰度区域；(b)边缘附近的细节，显示了一个水平灰度剖面及其一阶导数和二阶导数

例 10.4 带噪边缘的一阶导数和二阶导数的性质。

整个图像段的导数计算将在下一节中讨论。现在，我们的目的只是分析灰度剖面。

图 10.8 中的边缘是无噪声的。图 10.11 中第一列图像段是 4 个斜坡边缘的特写，这些边缘从左边的黑色区域到右边的白色区域过渡(记住，从黑到白的整个过渡是一个单一的边缘)。左上方的图像段无噪声。第一列中的其他三幅图像被均值为零、标准差分别为 0.1, 1.0 和 10.0 个灰度级的加性高斯噪声污染。每幅图像下面的图形是通过图像中心的水平灰度剖面线。所有图像具有 8 比特的灰度分辨率，并用 0 和 255 分别表示黑色与白色。

考虑中间一列顶部的图像。就像我们对图10.10(b)的讨论那样，左侧扫描线的导数在灰度恒定区域为零。这是在导数图像中显示的两个黑色条带。在斜坡上的各点处的导数是恒定的，并等于斜坡的斜率。在导数图像中这些恒定值显示为灰色。当我们沿中间列向下移动时，导数会变得与无噪声情形时越来越不相同。实际上，将中间列中最后一个剖面与斜坡边缘的一阶导数联系起来将会很困难。造成这种有意思的结果的原因是，左列图像中的噪声几乎不可见。这些例子很好地说明了导数对噪声的敏感性。

如预想的那样，二阶导数对于噪声甚至更为敏感。无噪声图像的二阶导数显示在右列的上部。白色和黑色的细垂直线是二阶导数的正分量和负分量，就像图 10.10 中解释的那样。这些图像中的灰色表示零(如先前讨论的那样，比例缩放导致零显示为灰色)。唯一的一个类似于无噪声情况的有噪声二阶导数图像对应于标准差为 0.1 的噪声。另两幅二阶导数图像和剖面清楚地表明了检测这些图像中的正分量和负分量的确很困难，而这些分量在边缘检测中确实是非常有用的二阶导数特性。

微弱的可见噪声严重影响检测边缘所用的两个关键导数这一事实，是我们应记住的一个重要问题。特别地，在类似于我们刚刚讨论的水平噪声很可能存在的应用中，使用导数之前应进行图像的平滑处理。

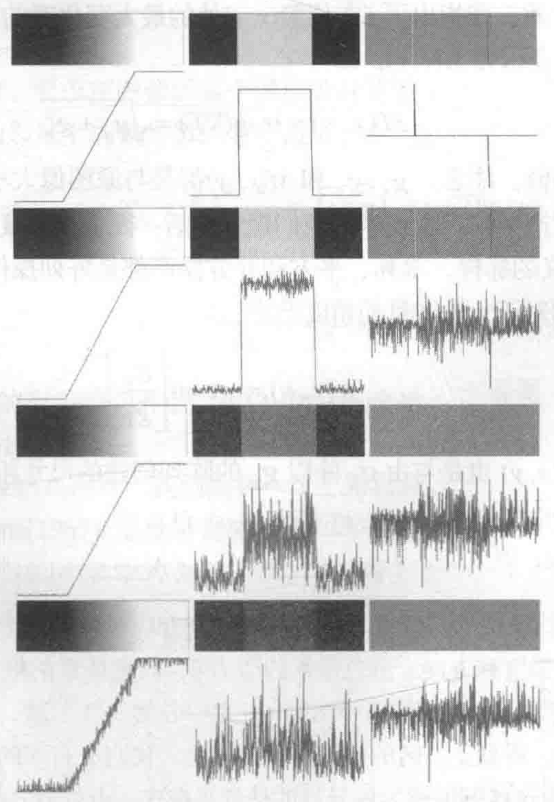


图 10.11 第一列：被均值为零、标准差分别为 0.0, 0.1, 1.0 和 10.0 个灰度级的随机高斯噪声所污染的斜坡边缘的图像和灰度剖面。第二列：一阶导数图像和灰度剖面线。第三列：二阶导数图像和灰度剖面线

726 ~ 727

我们用上面的解释给出执行边缘检测的三个基本步骤：

1. 为降噪对图像进行平滑处理。需要这一步的原因已由图10.11的第二列和第三列中的结果做了详细说明。
2. 边缘点的检测。如先前提及的那样，这是一个局部操作，从一幅图像中提取所有的点，这些点是变为边缘点的潜在候选者。
3. 边缘定位。这一步的目的是从候选边缘点中选择组成边缘点集合的真实成员。

本节剩下的内容探讨实现这些目标的一些技术。

10.2.5 基本边缘检测

如前一节说明的那样，为了达到寻找边缘的目的，检测灰度变化可用一阶或二阶导数来完成。本节讨论一阶导数，二阶导数将在10.2.6节中讨论。

图像梯度及其性质

为方便起见，这里重复给出3.6.4节中的一些公式。

为了在一幅图像f的(x, y)位置处寻找边缘的强度和方向，所选择的工具就是梯度，梯度用∇f来表示，并用向量来定义：

$$\nabla f \equiv \text{grad}(f) \equiv \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} \tag{10.2-9}$$

该向量有一个重要的几何性质，即指出了f在位置(x, y)处的最大变化率的方向。

向量∇f的大小(长度)表示为$M(x, y)$，即

$$M(x, y) = \text{mag}(\nabla f) = \sqrt{g_x^2 + g_y^2} \tag{10.2-10}$$

它是梯度向量方向变化率的值。注意，g_x, g_y和$M(x, y)$都是与原图像大小相同的图像，是x和y在f中的所有像素位置上变化时产生的。实践中，我们通常称后一图像为梯度图像，或在含义很清楚时简称为梯度。如2.6.1节中定义的那样，求和、平方和开方操作都是阵列操作。

梯度向量的方向由下列对于x轴度量的角度给出：

$$\alpha(x, y) = \arctan\left[\frac{g_y}{g_x}\right] \tag{10.2-11}$$

如梯度图像的情况那样，$\alpha(x, y)$也是与由g_y除以g_x的阵列创建的尺寸相同的图像。任意点(x, y)处一个边缘的方向与该点处梯度向量的方向$\alpha(x, y)$正交。

例10.5 梯度的性质。

图10.12显示了直边缘线段的局部放大。所显示的每个方块对应于一个像素，我们的目的是得到用一个方框强调的点处边缘的强度和方向。灰色像素的值为0，白色像素的值为1。下面我们说明该例子。为计算x方向和y方向的梯度，该方法使用一个以一点为中心的3×3邻域，简单地从底部一行的像素中减去顶部一行邻域中的像素，得到x方向的偏导数。类似地，我们从右列的像素中减去左列的像素得到y方向的偏导数。接下来，用这些差值作为偏导数的估计，在这一点处有$\partial f / \partial x = -2$和$\partial f / \partial y = 2$。从而有

726~728

$$\nabla f = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

由此，我们可以得到这一点处的 $M(x, y) = 2\sqrt{2}$ 。类似地，相同点处梯度向量的方向满足式(10.2-11)：$\alpha(x,y) = \arctan(g_y/g_x) = -45°$ ，它与相对于 x 轴的正方向度量的 135°相同。图 10.12(b)显示了该梯度向量及其方向角。

回忆 2.4.2 节可知，图像坐标系统的原点位于左上角，正 x 轴向下延伸，正 y 轴向右延伸。

图 10.12(c)说明了之前提到的一个重要事实，即某点的边缘与该点的梯度向量正交。因此，在这个例子中，边缘的方向角是 $\alpha - 90° = 45°$ 。图 10.12(a)中的所有边缘点都有相同的梯度，所以整个边缘段都处在相同的方向上。梯度向量有时也称为边缘法线。当向量通过除以其幅值［见式(10.2-10)］而归一化为单位长度时，结果向量通常称为边缘单位法线。

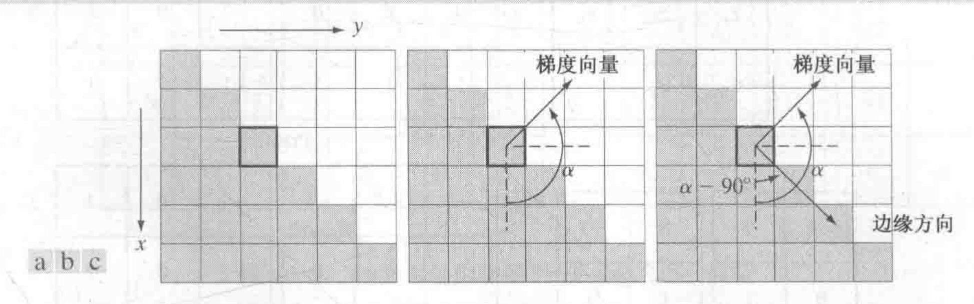


图 10.12　用梯度确定某个点处的边缘强度和方向。注意，某点处的边缘垂直于该点处的梯度向量的方向。图中的每个方块表示一个像素

梯度算子

要得到一幅图像的梯度，要求在图像的每个像素处计算偏导数 $\partial f/\partial x$ 和 $\partial f/\partial y$。我们处理的是数字量，因此要求关于一点的邻域上的偏导数的数字近似。由 10.2.1 节，我们可知

$$g_x = \frac{\partial f(x, y)}{\partial x} = f(x+1, y) - f(x, y) \tag{10.2-12}$$

和

$$g_y = \frac{\partial f(x, y)}{\partial y} = f(x, y+1) - f(x, y) \tag{10.2-13}$$

这两个公式对所有 x 和 y 的有关值可用图 10.13 中的一维模板通过对 $f(x, y)$的滤波来执行。

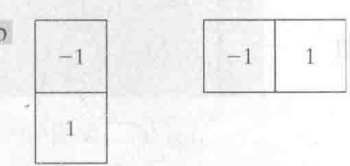


图 10.13　用于实现式(10.2-12)和式(10.2-13)的一维模板

当对对角线方向的边缘感兴趣时，我们需要一个二维模板。罗伯特交叉梯度算子(Roberts[1965])是最早尝试使用具有对角优势的二维模板之一。考虑图10.14(a)中的 3×3 区域。罗伯特算子以求对角像素之差为基础：

$$g_x = \frac{\partial f}{\partial x} = (z_9 - z_5) \tag{10.2-14}$$

和

$$g_y = \frac{\partial f}{\partial y} = (z_8 - z_6) \tag{10.2-15}$$

用于计算梯度偏导数的滤波器模板，通常称为梯度算子、差分算子、边缘算子或边缘检测子。

这些导数可以使用图 10.14(b)和(c)中的模板对图像进行滤波来实现。

729
~
730

2×2 大小的模板在概念上很简单，但它们对于用关于中心点对称的模板来计算边缘方向不是很有用，其最小模板大小为 3×3。这些模板考虑了中心点对端数据的性质，并携带有关于边缘方向的更多信息。用大小为 3×3 的模板来近似偏导数的最简数字近似由下式给出：

$$g_x = \frac{\partial f}{\partial x} = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3) \tag{10.2-16}$$

和

$$g_y = \frac{\partial f}{\partial y} = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7) \tag{10.2-17}$$

尽管这些公式包含了较大的邻域，但我们仍处理灰度值间的差，因此之前讨论一阶导数时得到的结论仍然适用。

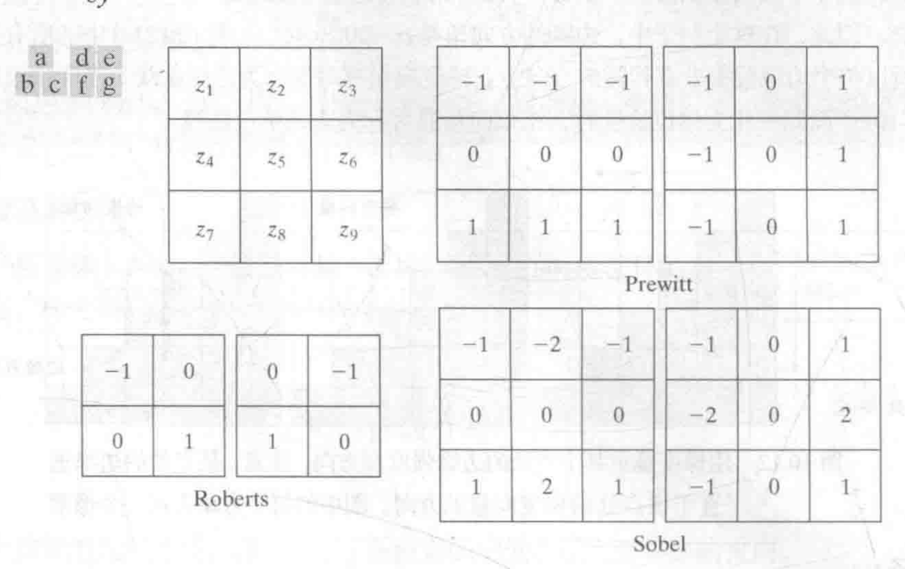


图 10.14　一幅图像的 3×3 区域(z 项是灰度值)和用于计算标记点 z_5 处的梯度的不同模板

在这些公式中，3×3 区域的第三行和第一行之差近似为 x 方向的导数，第三列和第一列之差近似为 y 方向的导数。直观上，我们可以预料这些近似要比用罗伯特算子得到的近似更准确。式(10.2-16)和式(10.2-17)可用图 10.14(d)和(e)中的两个模板通过滤波整个图像来实现。这两个模板称为 Prewitt 算子(Prewitt[1970])。

这两个公式的微小变化是，在中心系数上使用一个权值 2：

$$g_x = \frac{\partial f}{\partial x} = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3) \tag{10.2-18}$$

和

$$g_y = \frac{\partial f}{\partial y} = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7) \tag{10.2-19}$$

可以证明，在中心位置处使用 2 可以平滑图像(见习题 10.10)。图 10.14(f)和(g)显示了用于实现式(10.2-18)和式(10.2-19)的模板。这些模板称为 Sobel 算子(Sobel[1970])。

Prewitt 模板实现起来比 Sobel 模板更为简单，但它们之间计算上的小差别并不是问题。如前一节提到的那样，Sobel 模板能较好地抑制(平滑)噪声的特性这一事实使得它们更为可取，因为在处理导数时噪声抑制是一个重要的问题。注意，图 10.14 中所有模板中的系数之和为零，正如导数算子所预示的那样，恒定灰度区域的响应为零。

731 ~ 732

刚才讨论的模板用于在图像的每个像素位置处得到梯度分量 g_x 和 g_y。然后，用这两个偏导数估计边缘的强度和方向。计算梯度的幅度时，要求按式(10.2-10)所示的方法联合使用 g_x 和 g_y。但

这种实现并不总是令人满意，因为平方和平方根需要大量的计算开销。经常使用的一种方法是，用绝对值来近似梯度的幅值：

$$M(x, y) \approx |g_x| + |g_y| \tag{10.2-20}$$

该式不仅在计算上更有吸引力，而且仍保持着灰度级的相对变化。为这一优点付出的代价是，滤波器一般将不再是各向同性(旋转不变)的。然而，在使用 Prewitt 和 Sobel 这样的模板来计算 g_x 和 g_y 时，这并不是问题，因为这些模板仅对垂直和水平边缘才会给出各向同性的结果。不管使用这两个公式中的哪个，只在这两个方向上的边缘，结果才是各向同性的。此外，使用 Sobel 或 Prewitt 模板时，式(10.2-10)和式(10.2-20)对垂直和水平边缘给出相同的结果(见习题 10.8)。

可以修改图 10.14 中的 3×3 模板，以便它们沿对角线方向有最大的响应。图 10.15 显示了另外两个用于检测对角线方向边缘的 Prewitt 和 Sobel 模板。

a b c d

Prewitt (a)

0	1	1
−1	0	1
−1	−1	0

Prewitt (b)

−1	−1	0
−1	0	1
0	1	1

Sobel (c)

0	1	2
−1	0	1
−2	−1	0

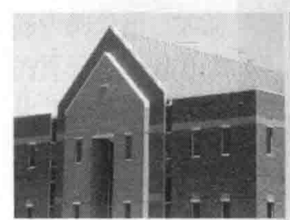
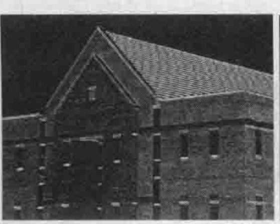
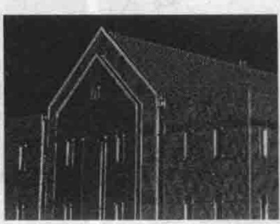
Sobel (d)

−2	−1	0
−1	0	1
0	1	2

图 10.15　检测对角边缘的 Prewitt 和 Sobel 模板

例 10.6　二维梯度幅值和角度的说明。

图 10.16 说明了两个梯度分量$|g_x|$和$|g_y|$的绝对值响应，以及由这两个分量之和形成的梯度图像。梯度的水平和垂直分量的方向性在图 10.16(b)和(c)中很明显。注意，图 10.16(b)中屋顶的瓦片、砖块的水平接缝和窗户的水平线段比比其他边缘强得多。相比之下，图 10.16(c)中正面和窗户的垂直分量更强。当一幅图像涉及的主要特征是边缘时，如梯度幅度图像，通常使用边缘图这一术语。在图 10.16(a)中，图像的灰度被标定在区间[0, 1]内。在本节讨论的各种边缘检测方法中，我们使用这一区间内的值来简化参数的选择。


a b c d

图 10.16　(a)灰度值已标定为范围[0, 1]内的、大小为 834×1114 的原图像；(b) $|g_x|$，用图 10.14(f)中的 Sobel 模板滤波图像得到的 x 方向上的梯度分量；(c) $|g_y|$，用图 10.14(g)中的模板得到的 y 方向上的梯度分量；(d)梯度图像$|g_x| + |g_y|$

图 10.17 显示了使用式(10.2-11)计算得到的梯度角度图像。通常，对于边缘检测而言，角度图像不像幅度图像那样有用，但它们可以作为用梯度幅值从图像中提取的信息的补充。例如，图 10.16(a)中的恒定灰度区域，诸如斜屋顶的前边缘和前墙顶部的水平条带，在图 10.17 中都是恒定的，这说明在这些区域中的所有像素位置处的梯度向量方向是相同的。如我们在 10.2.6 节说明的那样，在坎尼边缘检测算法的实现中，角度信息起着重要的支撑作用，坎尼边缘检测算法是在本章中我们讨论的最先进的边缘检测方法。

732 ~ 734

图10.16(a)中所示原图像的分辨率相当高(834×1114 像素)，且在图像获取的距离上，墙砖对于图像细节的贡献仍十分显著。这种精细的细节在边缘检测中通常是不符合需要的，因为它往往表现为噪声，导数计算会增强这种噪声，从而使图像中主要边缘的检测变得复杂。减少精细细节的一种方法是对图像进行平滑处理。图 10.18 显示了与图 10.16 中相同的图像序列，但先使用一个大小为 5×5 的均值滤波器对原图像进行了平滑(关于平滑滤波器，见 3.5 节)。现在每个模板的响应几乎未显示出由砖块造成的影响，得到的结果几乎都是主要的边缘。

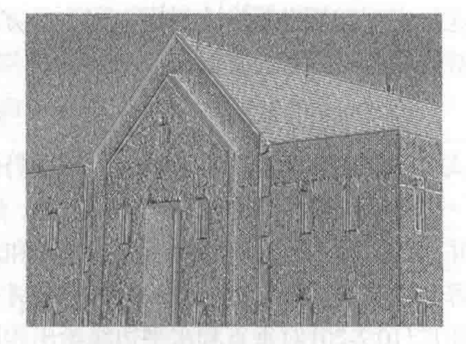


图 10.17　使用式(10.2-11)计算的梯度角度图像。在该图像中的恒定灰度区域指出这些区域中所有像素位置处的梯度向量的方向均相同

在图10.16 和图10.18 中，水平和垂直 Sobel 模板不区分 ±45°方向的边缘是很明显的。若沿对角方向强调边缘很重要，则应使用图10.15 中所示的一个模板。图10.19(a)和(b)分别显示了45°和 –45°的 Sobel 模板的绝对响应。在这些图中，这些模板的较强对角响应很明显。两个对角模板对水平边缘和垂直边缘都有类似的响应，但如期望的那样，它们在这些方向上的响应要比之前讨论的水平和垂直模板的响应弱。

平滑后图像的最大边缘强度(幅度)会随着平滑模板尺寸的增加而减小(见习题 10.13)。

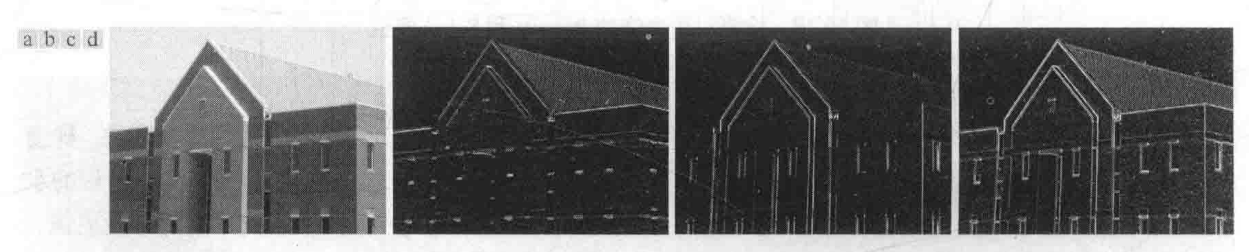


图 10.18　与图 10.16 中的相同图像序列，但在边缘检测之前，用大小为 5×5 的均值滤波器对原图像进行了平滑



图 10.19　对角边缘检测：(a)使用图 10.15(c)中的模板得到的结果；(b)使用图 10.15(d)中的模板得到的结果。两种情况下的输入图像都是图 10.18(a)

与阈值处理相结合的梯度

图 10.18 中的结果说明，在计算梯度前对图像进行平滑处理，边缘检测可做更多的选择。实现相同基本目标的另一种方法是，对梯度图像进行阈值处理。例如，图 10.20(a)显示了图 10.16(d)经阈值处理后的梯度图像，此时其值大于等于梯度图像的最大值的 33% 的像素显示为白色，而低于该阈值的像素则显示为黑色。将该图像与图 10.18(d)进行比较，我们看到阈值处理后的图像中边缘更少，

选择用于生成图 10.20(a)的阈值的目的，是消除砖块导致的多数较小边缘。回忆可知，这是在计算梯度前平滑图 10.16 中的图像的基本目标。

734～735

并且这幅图像中的边缘要清晰得多(例如，见屋瓦中的边缘)。另一方面，在阈值处理后的图像中，许多边缘（如屋顶边缘的 45°线）被断开了。

当目的是突出主要边缘并尽可能保持连接时，实践中通常既做平滑处理也做阈值处理。图 10.20(b)显示了图 10.18(d)经阈值处理后的结果，它是平滑后图像的梯度。结果中断开的边缘明显减少；例如，比较图 10.20(a)和(b)中的 45°边缘。当然，这些由于模糊处理其灰度值严重降低的边缘(即屋瓦中的边缘)可能因阈值处理而完全消除。在 10.2.7 节我们再回来讨论边缘断线的问题。

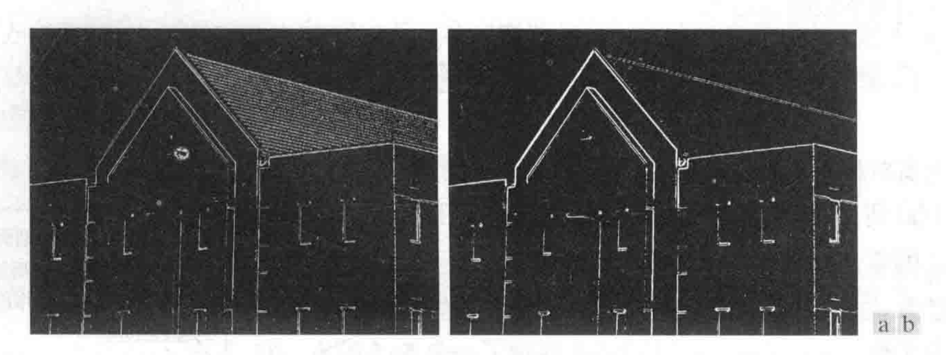


图 10.20　(a)图10.16(d)中图像经阈值处理后的图像，阈值选择为图像中最高值的 33%；该阈值刚好高到足以消除梯度图像中的多数砖块边缘；(b)图 10.18(d)中图像经阈值处理后的图像，它是使用等于图像中最高值的 33%的阈值得到的

10.2.6　更先进的边缘检测技术

前一节中讨论的边缘检测方法是以用一个或多个模板对图像进行滤波为基础的，而未对图像特性和噪声内容采取预防措施。本节讨论更为先进的技术，例如以图像噪声和边缘本身特性为基础边缘检测方法。

Marr-Hildreth 边缘检测器

Marr and Hildreth[1980]最早成功尝试了在边缘检测处理中结合更高级的分析。如前节所讨论的那样，当时，边缘检测方法是以使用较小的算子为基础的(如 Sobel 模板)。Marr and Hildreth 证明了：(1)灰度变化与图像尺寸无关，因此它们的检测要求使用不同尺寸的算子；(2)灰度的突然变化会在一阶导数中导致波峰或波谷，或在二阶导数中等效地引起零交叉(就像我们在图 10.10 中见到的那样)。

> 为证明边缘检测与图像尺寸无关，可以考虑图 10.8(c)中的屋顶边缘。如果图像的尺寸减小，则边缘会更细。

这些概念表明，用于边缘检测的算子应有两个明显的特点。第一个也最重要的特点是，它应该是一个能计算图像中每一点处的一阶导数或二阶导数的数字近似的微分算子。第二，它应能被“调整”以便在任何期望的尺寸上起作用，因此大算子也可用于检测模糊边缘，小算子可用于检测锐度集中的精细细节。

> 式(10.2-21)与二维高斯概率密度函数的定义的不同在于常数项 $1/2\pi\sigma^2$ 。如果在某个给定应用中期望得到一个精确的表达式，则乘以常数可添加到式(10.2-23)的最终结果中。

Marr and Hildreth 证明，满足这些条件的最令人满意的算子是滤波器 $\nabla^2 G$ ，如在 3.6.2 节中定义的那样，∇^2 是拉普拉斯算子 $(\partial^2/\partial x^2 + \partial^2/\partial y^2)$ ，而 G 是标准差为 σ (有时 σ 也称为空间常数)的二维高斯函数

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{10.2-21}$$

735～736

为求 $\nabla^2 G$ 的表达式，我们执行如下微分：

$$\nabla^2 G(x,y)=\frac{\partial^2 G(x,y)}{\partial x^2}+\frac{\partial^2 G(x,y)}{\partial y^2}=\frac{\partial}{\partial x}\left[\frac{-x}{\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}\right]+\frac{\partial}{\partial y}\left[\frac{-y}{\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}\right]$$
$$=\left[\frac{x^2}{\sigma^4}-\frac{1}{\sigma^2}\right]\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}}+\left[\frac{y^2}{\sigma^4}-\frac{1}{\sigma^2}\right]\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{10.2-22}$$

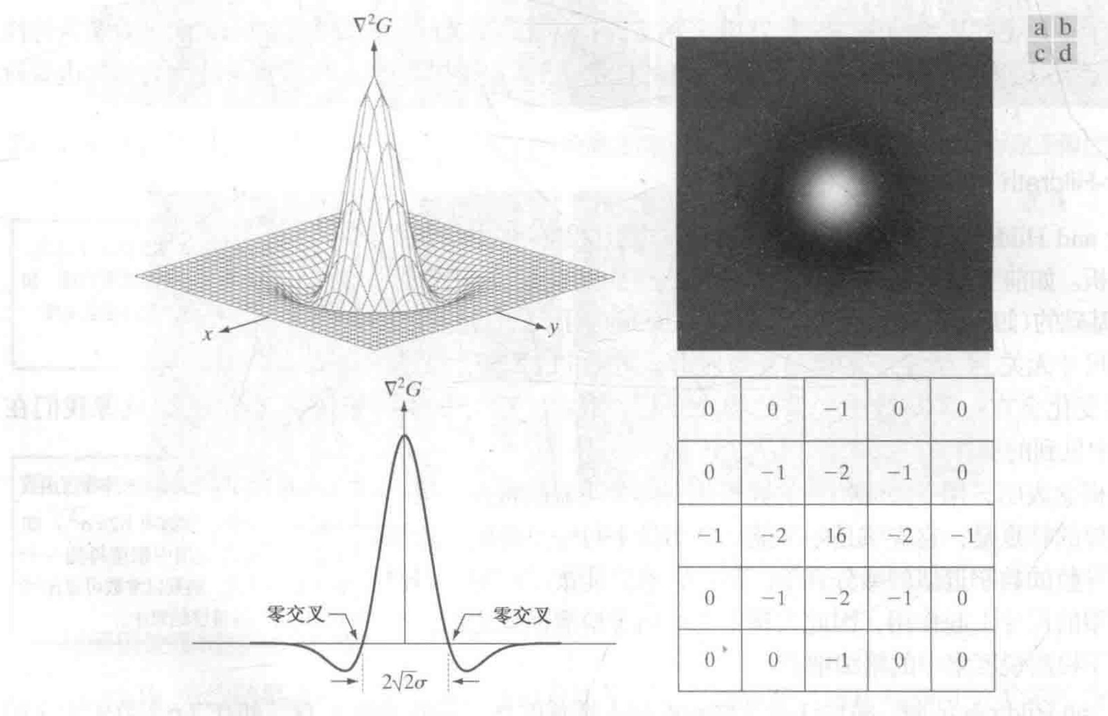
整理各项后给出如下最终表达式：

$$\nabla^2 G(x,y)=\left[\frac{x^2+y^2-2\sigma^2}{\sigma^4}\right]\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{10.2-23}$$

该表达式称为高斯拉普拉斯(LoG)。

> 注意图 10.21(c)中剖面与图 4.37(c)中高通滤波器之间的相似性。这样，我们可期望 LoG 的作用类似于一个高通滤波器。

图 10.21(a)到(c)显示了一个 LoG 的负函数的三维图、图像和剖面(注意 LoG 的零交叉出现在 $x^2+y^2=2\sigma^2$ 处，它定义了一个中心位于原点、半径为 $\sqrt{2}\sigma$ 的圆)。因为图 10.21(a)中显示的形状，LoG 函数有时也称为墨西哥草帽算子。图 10.21(d)显示了一个 5×5 模板，它近似于图 10.21(a)中的形状(在实践中，我们会用该模板的负模板)。这个近似并不是唯一的，其目的是获取 LoG 函数的基本形状；根据图 10.21(a)，这意味着一个正的中心项由紧邻的负区域包围，中心项的值以距原点的距离为函数而增大，而外层区域的值为零。系数之和必须为零，因此模板的响应在恒定灰度区域为零。



0	0	−1	0	0
0	−1	−2	−1	0
−1	−2	16	−2	−1
0	−1	−2	−1	0
0	0	−1	0	0

图 10.21 (a)负 LoG 的三维图；(b)显示为一幅图像的负 LoG 图；(c)显示了零交叉的图(a)的横截面；(d)对图(a)中形状 5×5 模板的近似，实际中将使用该模板的负值

736~737

任意尺寸的模板都可通过对式(10.2-23)取样并标定系数以使系数之和为零来生成。生成 LoG 滤波器的一种更有效的方法是，以希望的 $n\times n$ 尺寸对式(10.2-21)取样，然后将结果阵列与一个拉普拉斯模

板进行卷积①，例如图 10.4(a)中的模板。因为用一个系数之和为零的模板对图像阵列卷积产生一个元素之和也为零的结果(见习题 3.16 和习题 10.14)，故这种方法自动满足 LoG 滤波器系数之和为零的要求。在本节的后面将讨论 LoG 滤波器尺寸的选取问题。

选择算子 $\nabla^2 G$ 时要了解两个基本概念。第一，算子的高斯部分会模糊图像，从而在尺寸上将结构的灰度(包括噪声)降低到远小于 σ 的程度。与 3.5 节中讨论并在图 10.18 中使用的平均形式不同的是，高斯函数在空间和频率两个域平滑图像(见 4.8.3 节)，因而在原图像中引入不存在的人为干扰(如振铃)的可能性很小。另一个概念涉及 ∇^2，即滤波器的二阶导数部分。尽管一阶导数可用于检测灰度突变，但它们是有方向的算子。另一方面，拉普拉斯有各向同性(旋转不变)的重要优点，这不仅符合人的视觉系统特性(Marr[1982])，而且对任何模板方向的灰度变化有相等的响应，从而避免了使用多个模板去计算图像中任何点处的最强响应。

Marr-Hildreth 算法由 LoG 滤波器与一幅输入图像 $f(x, y)$卷积组成，即

$$g(x, y) = [\nabla^2 G(x, y)] \bigstar f(x, y) \tag{10.2-24}$$

然后寻找 $g(x, y)$的零交叉来确定 $f(x, y)$中边缘的位置。因为这些都是线性操作，故式(10.2-24)也可写为

$$g(x, y) = \nabla^2 [G(x, y) \bigstar f(x, y)] \tag{10.2-25}$$

它指出我们可以先使用一个高斯滤波器平滑图像，然后计算该结果的拉普拉斯。这两个公式给出了相同的结果。

> 该表达式是使用式(3.4-2)在空间域中实现的。也可使用式(4.7-1)在频率域中实现。

Marr-Hildreth 边缘检测算法可小结如下：

1. 用对式(10.2-21)取样得到的一个 $n\times n$ 高斯低通滤波器对输入图像滤波。
2. 计算由第一步得到的图像的拉普拉斯，例如，使用图10.4(a)的3×3 模板［第一步和第二步实现式(10.2-25)］。
3. 找到步骤 2 所得图像的零交叉。

为确定高斯滤波器的大小，仍采用位于二维高斯表面下其均值在 $\pm 3\sigma$ 之间的 99.7%。这样，作为一种经验法则，大小为 $n\times n$ 的 LoG 离散滤波器，其 n 值应是大于等于 6σ 的最小奇整数。n 小于该值的滤波器模板会“截断”LoG 函数，截断的程度与模板的大小成反比；而使用较大的模板对结果的影响不大。

在滤波后的图像 $g(x, y)$的任意像素 p 处，寻找零交叉的一种方法是，使用以 p 为中心的一个 3×3 邻域。p 点处的零交叉意味着至少有两个相对的邻域像素的符号不同。有 4 种要测试的情况：左/右、上/下和两个对角。如果 $g(x, y)$的值与一个阈值比较(一种常用方法)，那么不仅相对邻域的符号不同，而且它们的差值的绝对值还必须超过这个阈值。这时，我们称 p 为一个零交叉像素。下面的例 10.7 中将说明这一方法。

> 试图通过查找满足 $g(x, y) = 0$ 的坐标(x, y)来找到零交叉是不实际的，因为噪声和/或计算不准确。

零交叉是 Marr-Hildreth 边缘检测方法的关键特征。前段中讨论的方法很有吸引力，因为它实现简单，并且通常能给出较好的结果。如果在某个特殊应用中使用这一方法找到的零交叉位置的准确性不够，那么可采用由 Huertas and Medioni[1986]提出的采用亚像素精度来寻找零交叉的方法。

---

① LoG 是一个对称滤波器，所以使用相关或卷积的空间滤波将产生相同的结果。为与关于这一主题的文献保持一致，这里我们使用卷积这一术语来表示线性滤波。此外，也为读者提前了解这一术语，因为在其他内容中也会遇到它。记住，3.4.2 节末针对这一主题给出的注释很重要。

738 ~ 739

例 10.7 Marr-Hildreth 边缘检测算法说明。

图 10.22(a)显示了早先使用过的建筑物图像，图 10.22(b)是 Marr-Hildreth 算法的步骤 1 和步骤 2 得到的结果，其中 $\sigma = 4$（约为图像短边长度的 0.5%）和 $n=25$（如早先讨论的那样，大于等于 6σ 的最小奇整数）。如在图 10.5 中那样，这幅图像中的灰色调已被标定。图 10.22(c)显示了使用上面讨论过的阈值为零的 3×3 邻域方法得到的零交叉。注意，所有的边缘形成了一个闭环。这种所谓的意大利通心粉效应是使用零阈值方法的严重缺点（见习题 10.15）。我们使用正阈值避免闭环边缘。

图 10.22(d)显示了使用近似等于 LoG 图像最大值的 4%的阈值的结果。注意，很容易地检测出了大多数的主要边缘，并且滤掉了一些“无关”特征，例如砖块和屋瓦的边缘。如我们在下一节要说明的那样，使用前节讨论的基于梯度的边缘检测技术实际上是不可能得到这种类型的性能的。使用零交叉检测边缘的另一个重要结果是可得到 1 像素宽的边缘。这一特性简化了诸如边缘连接的后续阶段的处理。

a b c d

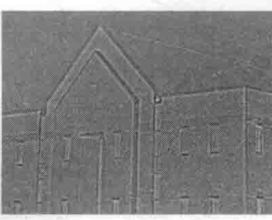
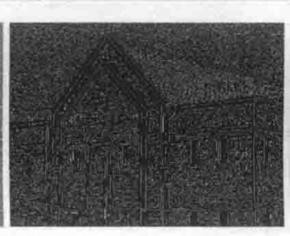





图 10.22 (a)大小为 834×1114 像素、灰度值已标定到范围[0, 1]的原图像；(b)Marr-Hildreth 算法的步骤 1 和步骤 2 得到的结果，其中 $\sigma = 4$ 和 $n = 25$；(c)使用零阈值的图(b)的零交叉(注意闭环边缘)；(d)使用等于图(b)中图像的最大值的 4%的阈值时，找到的零交叉。注意细边缘

考虑到早先提及的灰度变化取决于数值范围的事实，有时所用的过程是使用各种 σ 值来对一幅图像进行滤波。然后，所得零交叉边缘图与仅为全部图形保留的公共边缘相结合。这种方法可得到很有用的信息，但由于其复杂性，实践中它多被用做使用单一滤波器选择合适的 σ 值的设计工具。

Marr and Hildreth[1980]指出，使用高斯差分(DoG)来近似式(10.2-23)中的 LoG 滤波器是可能的：

$$\text{DoG}(x, y)=\frac{1}{2\pi\sigma_1^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma_1^2}}-\frac{1}{2\pi\sigma_2^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma_2^2}} \tag{10.2-26}$$

式中 $\sigma_1 > \sigma_2$。实验结果表明，在人的视觉系统中，某些“通道”就方向和频率而论是有选择性的，且可以使用式(10.2-26)以 1.75:1 的标准差比率来建模。Marr and Hildreth 建议，使用 1.6:1 的比率不仅可保持这些观察的基本特性，而且还可对 LoG 函数提供一个更接近的“工程”近似。为在 LoG 和 DoG 之间进行有意义的比较，对于 LoG，σ 值必须按照如下公式选择，以便 LoG 和 DoG 具有相同的零交叉（见习题 10.17）：

> 高斯差分是一个高通滤波器，详见 4.7.4 节中的讨论。

$$\sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 - \sigma_2^2}\ln\left[\frac{\sigma_1^2}{\sigma_2^2}\right] \tag{10.2-27}$$

使用这个 σ 值时，尽管 LoG 和 DoG 的零交叉相同，但它们的幅度大小会不同。我们可以通过标定这两个函数使得它们兼容，以便它们在原点处有相同的值。

图10.23(a)和(b)中的剖面线是分别使用标准差比率 1:1.75 和 1:1.6 产生的(按照惯例，所显示的曲线被反转，就像在图 10.21 中那样)。LoG 剖面线显示为实线，而 DoG 剖面线显示为虚线。所示曲线是通过 LoG 和 DoG 阵列中心的灰度剖面线，这两个阵列分别是通过对式(10.2-23)(前面有一个常数 $1/2\pi\sigma^2$)和式(10.2-26)取样产生的。所有曲线在原点处的幅度均被归一化为 1。如图 10.23(b)所示，1:1.6 的比率在 LoG 和 DoG 函数之间产生了一个更接近的近似。

739 ~ 741

LoG 和 DoG 滤波操作均可直接使用一维卷积代替二维卷积来实现（见习题 101.9）。对于一幅大小为 $M\times N$ 的图像和一个大小为 $n\times n$ 的滤波器，这样做可将每次卷积所需的乘法和加法次数，从二维卷积的与 n^2MN 成正比，减少到一维卷积的与 $n\ MN$ 成正比。这种实现上的差别是很有意义的。例如，若 $n=25$，则一维卷积实现的乘法与加法操作次数将比二维卷积实现的少 12 倍。

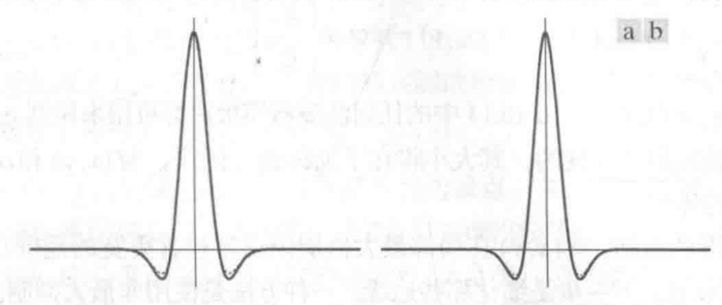


图10.23 (a)用比率为1.75 : 1的标准差的LOG(实线)和DOG(虚线)的负值剖面线，(b)使用 1.6 : 1 的比率得到的剖面线

坎尼边缘检测器

虽然其算法更为复杂，但本节讨论的坎尼检测器（Canny[1968]）是迄今为止讨论过的边缘检测器中最为优秀的。坎尼方法基于三个基本目标：

1. 低错误率。所有边缘都应被找到，并且应该没有伪响应。即检测到的边缘必须尽可能是真实的边缘。
2. 边缘点应被很好地定位。已定位边缘必须尽可能接近真实边缘。即由检测器标记为边缘的点和真实边缘的中心之间的距离应该最小。
3. 单一的边缘点响应。对于真实的边缘点，检测器仅应返回一个点。即真实边缘周围的局部最大数应该是最小的。这意味着在仅存一个单一边缘点的位置，检测器不应指出多个边缘像素。

坎尼的工作的本质是，从数学上表达前面的三个准则，并试图找到这些表达式的最佳解。通常，寻找一个满足前述目标的、严格形式的解是很困难的（或不可能的）。然而，对由加性高斯白噪声污染的一维台阶边缘使用数字最佳化会得到这样一个结论，即对最佳台阶边缘检测器的一个较好近似[①]是高斯一阶导数：

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}=\frac{-x}{\sigma^2}\mathrm{e}^{-\frac{x^2}{2\sigma^2}} \tag{10.2-28}$$

> 回忆可知，白噪声是指具有这样一个频谱的噪声，该频谱在一个指定的频率段中是连续且均匀的。白高斯噪声则是幅度值分布为高斯的白噪声。高斯白噪声是许多真实世界情形的较好近似，并且可产生数学上可处理的模型。高斯白噪声的一个有用性质是其值是统计独立的。

把该结果推广到二维情况会涉及识别一维方法是否仍适用于边缘法线的方向（见图 10.12）。因为预先法线方向未知，这就要求在所有可能的方向上应用一维边缘检测器。这一任务可由如下方法来近似：首先使用一个环形二维高斯函数平滑图像，计算结果的梯度，然后使用梯度幅度和方向来估计每一点处的边缘强度与方向。

令 $f(x,y)$表示输入图像，$G(x,y)$表示高斯函数：

$$G(x,y)=\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{10.2-29}$$

我们用 G 和 f 的卷积形成一幅平滑后的图像 $f_s(x,y)$：

$$f_s(x,y)=G(x,y)\bigstar f(x,y) \tag{10.2-30}$$

741 ~ 742

① Canny[1986]证明，使用高斯近似与使用最优数值解相比，所得结果仅差约 20%。这种幅度差在多数应用中通常是难以察觉的。

像 10.2.5 节讨论的那样，这一操作源自计算梯度幅度和方向：

$$M(x, y) = \sqrt{g_x^2 + g_y^2} \tag{10.2-31}$$

和

$$\alpha(x, y) = \arctan\left[\frac{g_y}{g_x}\right] \tag{10.2-32}$$

式中，$g_x = \partial f_s / \partial x$，$g_y = \partial f_s / \partial y$。图 10.14 中的任何滤波器模板对都可用来得到 g_x 和 g_y。式(10.2-30)是使用大小为 $n \times n$ 的高斯模板实现的，其大小将在下面讨论。记住，$M(x,y)$ 和 $\alpha(x,y)$ 是与计算出它们的图像的尺寸相同的阵列。

因为它是使用梯度产生的，$M(x,y)$ 在局部最大值周围通常包含更宽的范围(回忆10.2.1节中关于使用梯度得到边缘的讨论)。下一步是细化那些边缘。一种方法是使用非最大抑制。这可通过几种方式来实现，但该方法的本质是指定边缘法线的许多离散方向(梯度向量)。例如，在一个 3×3 区域内，对于一个通过该区域中心点的边缘，我们可以定义四个方向[①]：水平、垂直、+45°和−45°。图10.24(a)显示了一个水平边缘的两个可能方向。因为我们必须把所有可能的边缘方向量化为四个方向，故必须定义一个方向范围，在该范围内，我们考虑一个水平方向的边缘。我们由边缘法线的方向来确定边缘方向，边缘法线方向可以直接使用式(10.2-32)从图像数据得到。如图 10.24(b)所示，若边缘法线方向的范围是从−22.5°到 22.5°，或从−157.5°到 157.5°，则我们称该边缘为水平边缘。图 10.24(c)显示了对应于所考虑的四个方向的角度范围。

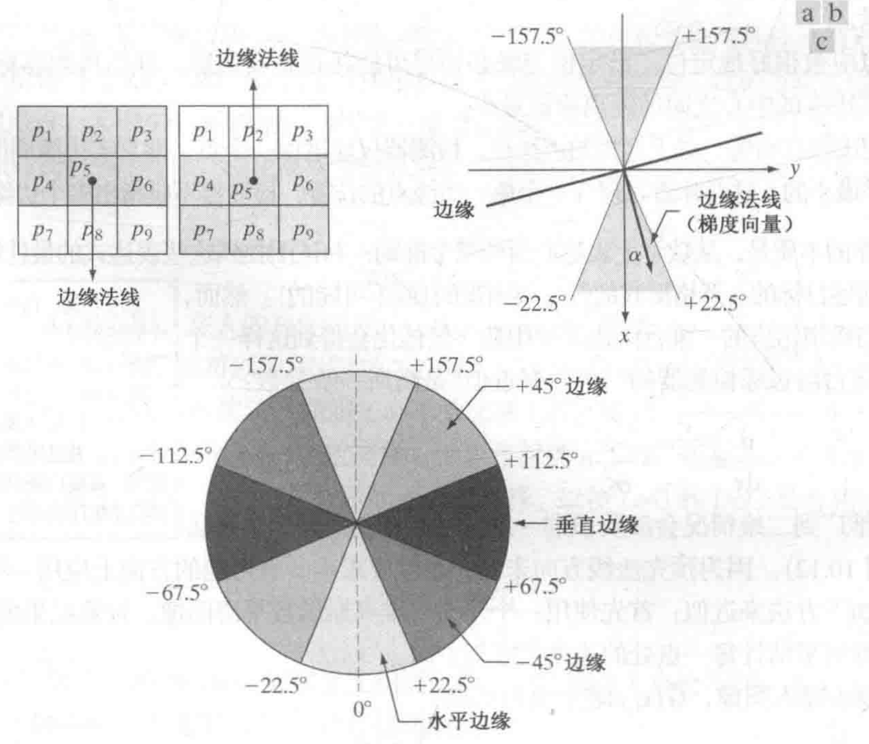


图 10.24 (a)在一个 3×3 邻域中，一个水平边缘(灰色)的两个可能方向；(b)一个水平边缘的边缘法线的方向角 α 的值的范围(灰色)；(c)在一个 3×3 邻域中，4 种类型的边缘方向的边缘法线的角度范围。每个边缘方向都有两个范围，以对应的灰色显示

742 ~ 743

① 记住，每个边缘均有两个可能的方向。例如，其法线为 0°的边缘和法线为 180°的边缘是同一个水平边缘。

令 d_1, d_2, d_3 和 d_4 表示刚才讨论的 3×3 区域的四个基本边缘方向：水平、$-45°$、垂直、$+45°$。对于 $\alpha(x,y)$ 中以每点 (x,y) 为中心的 3×3 区域，我们可给出如下非最大抑制方案：

1. 寻找最接近 $\alpha(x,y)$ 的方向 d_k。
2. 若 $M(x,y)$ 的值至少小于沿 d_k 的两个邻居之一，则令 $g_N(x,y)=0$（抑制）；否则，令 $g_N(x,y)=M(x,y)$，这里 $g_N(x,y)$ 是非最大抑制后的图像。例如，参考图10.24(a)，令 (x,y) 在 p_5 处，并假设一个水平边缘通过 p_5，在步骤 2 中我们感兴趣的像素是 p_2 和 p_8。图像 $g_N(x,y)$ 仅包含细化后的边缘；它等于抑制了非最大边缘点的 $M(x,y)$。

最后的操作是对 $g_N(x,y)$ 进行阈值处理，以便减少伪边缘点。在 10.2.5 节中，我们使用单阈值进行阈值处理，其中低于该阈值的所有值均置为零。若我们设的阈值过低，则仍会存在一些伪边缘（称为伪正值）。若阈值设得过高，则会删除实际上有效的边缘点（伪负值）。如我们在 10.3.6 节中讨论的那样，坎尼算法通过使用滞后阈值试图改进这一状况，它使用两个阈值：一个低阈值 T_L 和一个高阈值 T_H。坎尼建议，高阈值和低阈值的比率应为 2∶1 或 3∶1。

我们可以将阈值操作想象为创建两幅附加的图像

$$g_{NH}(x, y) = g_N(x, y) \geqslant T_H \tag{10.2-33}$$

和

$$g_{NL}(x, y) = g_N(x, y) \geqslant T_L \tag{10.2-34}$$

其中，开始时 $g_{NH}(x,y)$ 和 $g_{NL}(x,y)$ 均被置为零。阈值处理后，$g_{NH}(x,y)$ 的非零像素通常比 $g_{NL}(x,y)$ 少，但 $g_{NH}(x,y)$ 中的所有非零像素都将包含在 $g_{NL}(x,y)$ 中，因为后一图像是使用一个低阈值形成的。通过令

$$g_{NL}(x, y) = g_{NL}(x, y) - g_{NH}(x, y) \tag{10.2-35}$$

我们从 $g_{NL}(x,y)$ 中删除所有来自 $g_{NH}(x,y)$ 的非零像素。$g_{NH}(x,y)$ 和 $g_{NL}(x,y)$ 中的非零像素可分别视为“强”和“弱”边缘像素。

阈值处理后，$g_{NH}(x,y)$ 中的所有强像素均被假设为有效的边缘像素，并被立即标记。取决于 T_H 的值，$g_{NH}(x,y)$ 中的边缘通常会存在缝隙。较长的边缘用下列步骤形成：

(a) 在 $g_{NH}(x,y)$ 中定位下一个未被访问的边缘像素 p。
(b) 在 $g_{NL}(x,y)$ 中将所有弱像素标记为有效边缘像素，用 8 连通的连接方法连接到 p。
(c) 若 $g_{NH}(x,y)$ 中的所有非零像素已被访问，则跳到步骤(d)，否则返回步骤(a)。
(d) 将 $g_{NL}(x,y)$ 中未标记为有效边缘像素的所有像素置零。

在这一过程的末尾，将来自 $g_{NL}(x,y)$ 的所有非零像素附加到 $g_{NH}(x,y)$，用坎尼算子形成最终的输出图像。

我们使用了两幅附加的图像 $g_{NH}(x,y)$ 和 $g_{NL}(x,y)$ 来简化讨论。在实际中，滞后阈值处理可在非最大抑制期间直接执行，并且可以通过形成一个与它们相联系的强像素和弱像素的列表在 $g_N(x,y)$ 中直接实现。

坎尼边缘检测算法的基本步骤总结如下：

1. 用一个高斯滤波器平滑输入图像。
2. 计算梯度幅值图像和角度图像。
3. 对梯度幅值图像应用非最大抑制。
4. 用双阈值处理和连接分析来检测并连接边缘。

743 ~ 745

尽管非最大抑制后的边缘比原来的梯度边缘细，但仍保持了比 1 像素粗的边缘。为得到 1 像素粗的边缘，通常要在步骤 4 后执行一次边缘细化算法(见 9.5.5 节)。

如先前提到的那样，平滑是用指定的大小为 $n\times n$ 的高斯模板与输入图像进行卷积来完成的。我们可以使用前面章节讨论过的与 Marr-Hildreth 算法有关的方法确定 n 值。即使用对式(10.2-29)取样所产生的一个滤波器模板，使 n 是大于等于 6σ 的最小奇整数，进而提高高斯滤波器的“完美”平滑能力。若实际考虑要求一个较小的滤波器模板，则折中方案是较小 n 值的平滑效果会差一些。

关于实现的最终意见如下：如先前讨论Marr-Hildreth 边缘检测器那样，式(10.2-29)中的二维高斯函数被分解为两个一维高斯函数的乘积。这样，坎尼算法的步骤 1 就可用一维卷积公式来表达，即一次对图像中的一行(列)进行操作，然后再对处理结果中的列(行)操作。此外，如果使用式(10.2-12)和式(10.2-13)中的近似，也可将步骤 2 中要求的梯度计算以一维卷积来实现(见习题 10.20)。

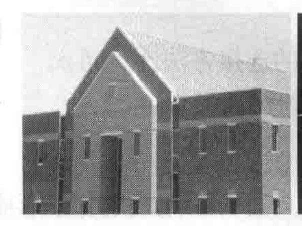
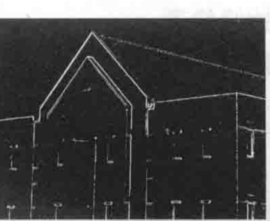
例 10.8　坎尼边缘检测方法说明。

图 10.25(a)显示了我们熟悉的建筑物图像。为了比较，图 10.25(b)和(c)分别显示了先前在图 10.20(b)中用阈值化后的梯度得到的结果，以及在图 10.22(d)中使用 Marr-Hildreth 检测子得到的结果。回忆可知，选取生成这两幅图像所用的参数来检测主要边缘，同时减少“无关”特征，如由砖块和屋瓦引起的边缘。

图 10.25(d)显示了用坎尼算法得到的结果，使用的参数为：$T_L = 0.04$，$T_H = 0.10$(低阈值的 2.5 倍)，$\sigma = 4$，以及对应于大于 6σ 的最小奇整数的、大小为 25×25 的模板。这些参数可交互式地选择，以达到前面章节中对梯度和 Marr-Hildreth 图像扫描规定的目标。将坎尼图像与其他两幅图像比较，我们看到主要边缘在细节上已有明显改进，同时在坎尼结果中抑制了更多的无关特性。例如，注意坎尼算法检测到了图像上部镶嵌在墙砖中间的水泥条带的两个边缘，然而，阈值处理后的梯度图像中则消除了这两个边缘，而 Marr-Hildreth 图像中仅包含上面的一个边缘。在滤除无关细节方面，坎尼图像确实不包含由屋瓦引起的单个边缘；在其他两幅图像中却不是这样。在坎尼图像中，关于连续性、细度和笔直度等线的质量也很出众。这种结果使得坎尼算法成为边缘检测的首选工具。

> 这里给出的几个阈值仅应在相对情形下考虑。多数算法实现会涉及各个标定步骤，如将图像的值域标定为范围[0, 1]。从这一例子可以看出，不同的标定方案明显会要求不同的阈值。

a b c d

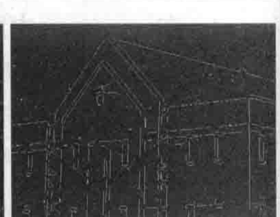


图 10.25　(a)大小为 834×1114 像素、灰度值标定为范围[0, 1]内的原图像；(b)平滑过的图像经阈值处理后得到的梯度图像；(c)使用 Marr-Hildreth 算法得到的图像；(d)使用坎尼算法得到的图像。注意，与其他两幅图像相比，坎尼图像有明显的改进

例 10.9　本节中讨论的三种主要边缘检测方法的另一说明。

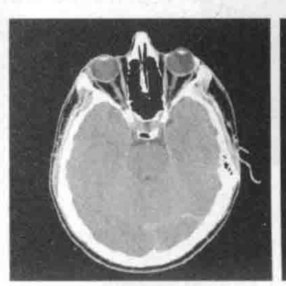
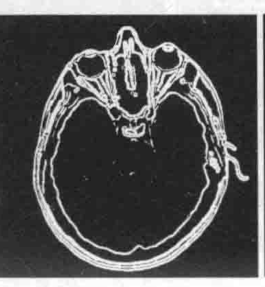
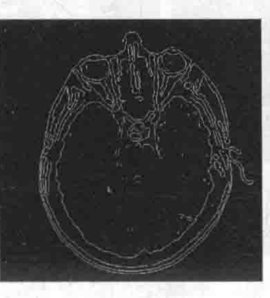
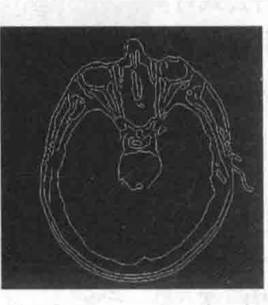
作为本节中讨论的三种主要边缘检测方法的另一比较，考虑图10.26(a)，它显示了一幅大小为 512×512 的头部 CT 图像。在这个例子中，我们的目的是提取大脑外轮廓(图像中的灰色区域)、脊髓区域轮廓和头部外轮廓的边缘。在消除眼部和大脑区域中与灰色有关的边缘细节时，我们希望生成尽可能最细且连续的轮廓。

745 ~ 746

图10.26(b)显示了一幅阈值处理后的梯度图像，它已先用一个 5×5 均值滤波器平滑过。达到所示结果

要求的阈值是该梯度图像的最大值的15%。图10.26(c)显示了使用Marr-Hildreth边缘检测算法得到的结果，所用的阈值为0.002，$\sigma=3$，模板大小为19×19像素。图10.26(d)是使用坎尼算法得到的，所用的$T_L=0.05$，$T_H=0.15$（低阈值的3倍），$\sigma=2$，模板大小为13×13，像Marr-Hildreth情况那样，它对应于大于6σ的最小奇整数。

在边缘质量及消除无关细节的能力方面，图10.26中的结果相当接近前一例子中的结果与结论。还要注意，坎尼算法是唯一能为大脑后边界产生完整无断裂边缘的过程，也是在消除原始图像中与灰度物质有关的所有边缘时，有可能发现最好轮廓的过程。

a b c d

图10.26　(a)大小为512×512像素、灰度值已标定到范围[0, 1]内的原始头部CT图像；(b)平滑后的图像经阈值处理后的梯度图像；(c)用Marr-Hildreth算法得到的图像；(d)用坎尼算法得到的图像(原图像由Vanderbilt大学的David R. Pickens博士提供)

正像预期的那样，坎尼算法改进性能所付出的代价是，它要比早些时候讨论的两种方法实现起来更复杂，同时执行时间更长。在某些应用中，如实时工业图像处理，成本和速度需求通常要求我们使用更简单的技术，主要是阈值梯度方法。当关注的主要是边缘质量时，通常会使用Marr-Hildreth算法和坎尼算法，特别是坎尼算法。

10.2.7　边缘连接和边界检测

理想情况下，边缘检测应该仅产生位于边缘上的像素集合。实际上，由于噪声、不均匀照明引起的边缘间断，以及其他引入灰度值虚假的不连续的影响，这些像素并不能完全描述边缘特性。因此，一般是在边缘检测后紧跟连接算法，将边缘像素组合成有意义的边缘或区域边界。本节讨论三种基本的边缘连接方法，它们是实际中使用的代表性技术。第一种方法需要有关局部区域中边缘点(如一个3×3邻域)的知识；第二种方法要求区域边界上的点已知；第三种技术是处理整个边缘图像的全局方法。

局部处理

连接边缘点最简单的方法之一是，在每个点(x, y)处的一个小邻域内分析像素的特点，该点是用前节讨论过的技术之一声明的边缘点。根据预定的准则，将所有的相似点连接起来，以形成根据指定准则满足相同特性像素的一条边缘。

在这种类型的分析中，用于确定边缘像素相似性的两个主要性质是：(1)梯度向量的强度(幅度)；(2)梯度向量的方向。第一个性质基于式(10.2-10)。令S_{xy}表示一幅图像中以点(x, y)为中心的一个邻域的坐标集合。如果

$$|M(s,t)-M(x,y)|\leqslant E \tag{10.2-36}$$

式中，E是一个正阈值。

梯度向量的方向角由式(10.2-11)给出。如果

746~748

$$|\alpha(s, t)-\alpha(x, y)| \leqslant A \tag{10.2-37}$$

式中，A 是一个正角度阈值，则 S_{xy} 中坐标(s, t)处的一个边缘像素有一个与(x, y)处像素类似的角度。如10.2.5 节中注释的那样，(x, y)处的边缘的方向垂直于该点处梯度向量的方向。

如果既满足幅度准则，也满足方向准则，则 S_{xy} 中，坐标为(s, t)的像素被连接到坐标为(x, y)的像素。在图像中的每个位置重复这一处理。当邻域的中心从一个像素移到另一个像素时，必须将已连接的点记录下来。简单的记录过程是对每组被连接的像素分配不同的灰度值。

前面的公式计算代价很高，因为必须检验每个点的所有邻点。一种特别适合于实时应用的简化由如下步骤组成：

1. 计算输入图像 $f(x,y)$ 的梯度幅度阵列 $M(x,y)$ 和梯度角度阵列 $\alpha(x,y)$ 。
2. 形成一幅二值图像 g，任何坐标对(x, y)处的值由下式给出：

$$g(x, y)=\begin{cases}1, & M(x,y)>T_M \text{且} \alpha(x, y)=A \pm T_A \\ 0, & \text{其他}\end{cases}$$

   式中，T_M 是一个阈值，A 是一个指定的角度方向，$\pm T_A$ 定义了一个关于 A 的可接受方向“带宽”。
3. 扫描 g 的行，并在不超过指定长度 K 的每一行中填充(置 1)所有缝隙(0 的集合)。注意，按照定义，缝隙一定要限制在一个 1 或多个 1 的两端。分别处理各行，它们之间没有记忆。
4. 在任何其他方向 θ 上检测缝隙，以该角度旋转 g，并应用步骤 3 中的水平扫描过程。然后，将结果以 $-\theta$ 旋转回来。

当目的在于水平边缘连接和垂直边缘连接时，步骤 4 就变为一个简单的过程，在该过程中，g 被旋转 90°，扫描各行后，结果再被旋转回来。这是实践中最常用的方法，如下例所示，这种方法可以产生很好的结果。通常，图像旋转是代价很高的计算处理，因此，在要求多角度方向上的连接时，把步骤 3 和步骤 4 组合成单个放射状扫描过程更为实用。

例 10.10 使用局部处理的边缘连接。

图 10.27(a)显示了一辆汽车尾部的图像。该例的目的是说明用前述算法来寻找大小适合车牌的矩形的应用。该矩形可以通过检测强的水平和垂直边缘构成。图 10.27(b)显示了梯度幅度图像 $M(x,y)$ ，图 10.27(c)和(d)显示了该算法步骤 3 和步骤 4 的结果，其中，令 T_M 等于最大梯度值的 30%， $A=90°$ ， $T_A=45°$ ，并填充了全部 25 个或更少像素的缝隙(约为图像宽度的 5%)。为检测车牌的全部拐角和汽车的后窗，要求使用一个较大范围的容许角度方向。图 10.27(e)是前两幅图像逻辑“或”(OR)操作的结果，图10.27(f)是使用 9.5.5 节讨论的细化过程细化图 10.27(e)得到的。如图 10.16(f)所示，在图像中清楚地检测到了对应于车牌的矩形。由于美国汽车牌照的宽高比是与众不同的 2:1，所以利用这一事实从图像的所有矩形中简单地分离出牌照很简单。

区域处理

通常，图像中感兴趣区域的位置已知或可以确定。这意味着在相应的边缘图像中关于像素的区域成员的知识是可用的。在这种情况下，我们可以使用在区域的基础上连接像素的技术，所期望的结果是该区域边界的近似。对这类处理的一种方法是函数近似，这里，我们对已知点拟合一条二维曲线。我们的兴趣通常在于快速执行的技术，它可对边界的基本特性产生一个近似，例如端点和凹点。多边形近似尤其有吸引力，因为在保持边界表示(如多边形的顶点)相对简单的情况下，它们可以捕捉基本形状特性。本节阐述和说明一种适合这一目的算法。

748~750

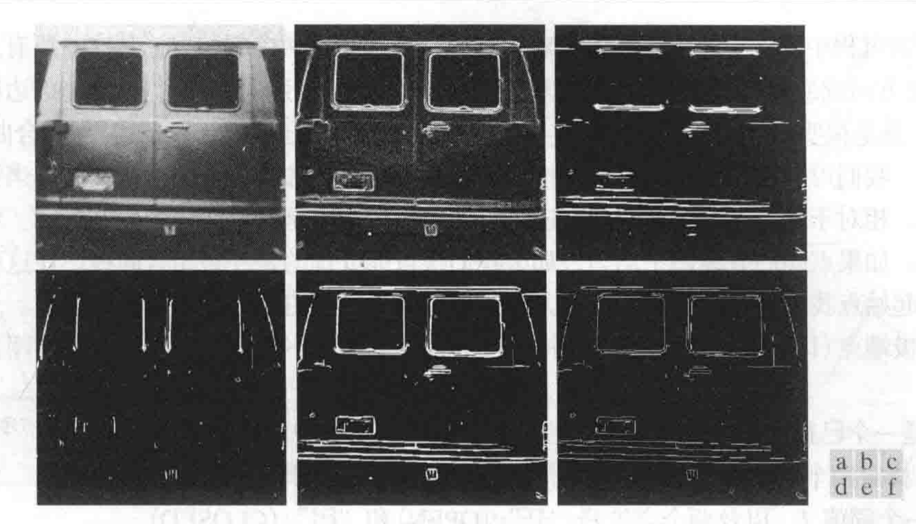


图 10.27 (a)一幅大小为 534×566 的汽车后部图像；(b)梯度幅度图像；(c)水平连接的边缘像素；(d)垂直连接的边缘像素；(e)前两幅图像的逻辑“或”(OR)；(f)用形态学细化得到的最终结果(原图像由 Perceptics Corporation 提供)

在说明该算法之前，我们用一个简单的例子来讨论该过程的机理。图10.28 显示了表示一条非闭合曲线点的集合，其中端点标记为 A 和 B。按照定义，这两个点是多边形的顶点。首先，我们计算通过 A 点和 B 点的一条直线的参数。然后，计算曲线上所有其他点到这条直线的垂直距离，并选择产生最大距离的点(目的是解决任意性)。若该距离超过了指定的阈值 T，则将标为 C 的相应点宣布为一个顶点，如图10.28(c)所示。然后，建立从 A 到 C 和从 C 到 B 的直线，得到从 A 至 C 间的所有点到线 AC 的距离。如果距离超过了 T，相应于最大距离的点被宣布为一个顶点 D；否则，对于该线段就没有新的顶点要宣布。类似的过程用于 C 和 B 之间的点。图10.28(b)显示了结果，图 10.28(c)显示了下一步的结果。继续这一迭代过程，直到没有点满足阈值测试为止。图10.28(d)显示了最终的结果，如我们看到的那样，该结果是拟合给定点的曲线形状的一种合理近似。

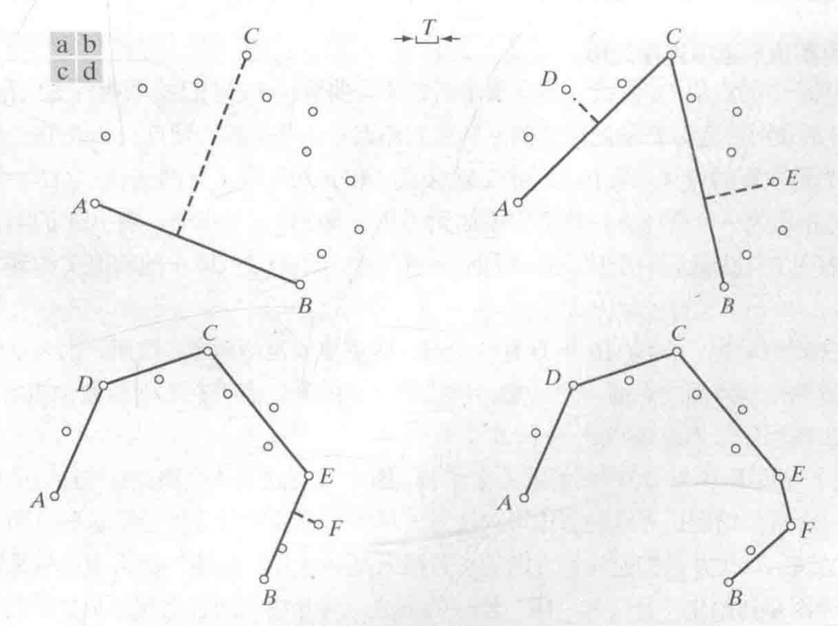


图 10.28 多边形拟合算法迭代说明

750 ~ 751

刚才解释的过程中隐含了两个重要的需求。首先，必须指定两个起始点；其次，所有点都必须排序(即按顺时针方向或逆时针方向)。任意一个二维点集未形成一个连通路径时(如典型的边缘图像中的情况)，这并不总是很明显，无论是点属于边界线段(开放曲线)，还是属于一条边界(闭合曲线)。给定已排序的点后，我们可以用分析点间的距离推断我们处理的是开放的还是闭合的曲线。当我们穿过这些点的序列时，相对于其他点的距离的两个连续点间较大的距离是开放曲线的较好指示。然后使用端点开始该过程。如果点间的分离趋于均匀，那么我们很可能处理的是一条闭合曲线。在这种情况下，对于选取两个起始点我们有几种选择。一种方法是选择点集中最右边的点和最左边的点。另一种方法是寻找曲线的极端点(11.2.1 节中讨论了一种这样的方法)。寻找一个多边形拟合开放和闭合曲线的算法描述如下：

> 关于创建已排序点序列的算法，见 11.1.1 节。

1. 令 P 是一个已排序序列，显然，这些点是一幅二值图像中的 1 值点。指定两个起始点 A 和 B。它们是多边形的两个起始顶点。
2. 指定一个阈值 T，以及两个空堆栈“开”(OPEN)和“闭”(CLOSED)。
3. 若 P 中的点对应于一条闭合曲线，则把 A 放到“开”中，并把 B 放到“开”和“闭”中。若点对应于一条开放曲线，则把 A 放到“开”中，而把 B 放到“闭”中。
4. 计算从“闭”中最后一个顶点到“开”中最后一个顶点的线的参数。
5. 计算步骤 4 所得的直线至 P 中所有点的距离，序列把它们放到步骤 4 所得的两个顶点之间。选择具有最大距离 D_{max} 的点 V_{max}(解决任意性问题)。
6. 若 $D_{max} > T$，则把 V_{max} 作为一个新顶点放在“开”的末尾。转到步骤 4。
7. 否则，从“开”中移除最后一个顶点，并把它作为“闭”的最后一个顶点插入。
8. 若“开”非空，转到步骤 4。
9. 否则，退出。“闭”中的顶点就是拟合 P 中的点的多边形的顶点。

> 为堆栈名称使用的 OPEN 和 CLOSED 与非闭合曲线和闭合曲线无关。堆栈名称仅指出一个存储最终(CLOSED)顶点或过渡期间(OPEN)的顶点的堆栈。

下面的两个例子说明了该算法的机理。

例 10.11　用多边形近似连接边缘。

考虑图 10.29(a)中的点集 P。假设这些点属于一条闭合曲线，且它们已按顺时针方向排序(注意有些点并不相邻)，A 和 B 分别被选为 P 中的最左侧点和最右侧点。这些是起始顶点，如表 10.1 所示。选取序列中的第一个点作为最左侧的点 A。图 10.29(b)仅显示了 A 和 B 间线段上方的点(标为 C)，该点满足算法的步骤 6，因此将它指定为一个新顶点，并将它添加到“开”堆栈内的顶点中。表 10.1 的第二行显示了被检测的 C，第三行表明它作为最后一个顶点被添加到“开”中。图 10.29(b)中的阈值 T 约等于网格图中的一格半。

注意，在图 10.29(b)中，直线 AB 下方有一个点，该点也满足步骤 6。然而，因为这些点已被排序，故一次只能检测这两个顶点间的点的一个子集。该线段下方的其他点将在后面检测，如图 10.29(e)所示。关键是总要按给出的顺序来追踪这些点。

表 10.1 说明了导致图 10.29(h)中的解的各个步骤。图中检测到了 4 个顶点，并且该图用直线段连接这 4 个顶点形成了一个近似给定边界点的多边形。注意，尽管是按顺时针方向追踪这些点而生成的顶点，但表中检测到顶点 B, C, A, D, B 是按逆时针方向的。若输入是一条开放曲线，那么顶点将是顺时针顺序的。导致这种差异的原因是初始化“开”和“闭”堆栈的方法。对开放和闭合曲线形成的堆栈“闭”的差别，也会导致闭合曲线中第一个顶点与最后一个顶点的重复。这与只给出顶点如何来区分开放和闭合多边形是一致的。

751 ~ 753

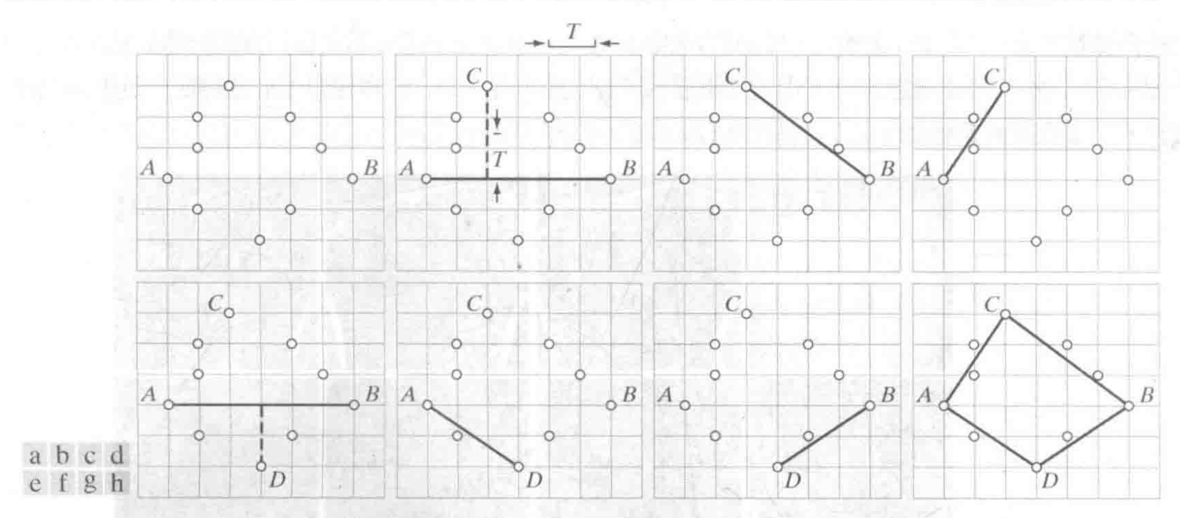


图 10.29 (a)一条顺时针路径中的点集(标为 A 和 B 的点被选为起始顶点)；(b)点 C 到过 A 和 B 的直线的距离是 A 和 B 间所有点中最大的，并且通过了阈值测试，因此 C 是一个新顶点；(d)~(g)算法的各个进程；(h)用直线连接形成一个多边形所显示的最后的顶点。表 10.1 说明了一步步的细节

表 10.1 例 10.11 机理的详细步骤

闭 合	非 闭 合	已处理的曲线段	生成的顶点
B	B, A	—	A, B
B	B, A	(BA)	C
B	B, A, C	(BC)	—
B, C	B, A	(CA)	—
B, C, A	B	(AB)	D
B, C, A	B, D	(AD)	—
B, C, A, D	B	(DB)	—
B, C, A, D, B	空	—	—

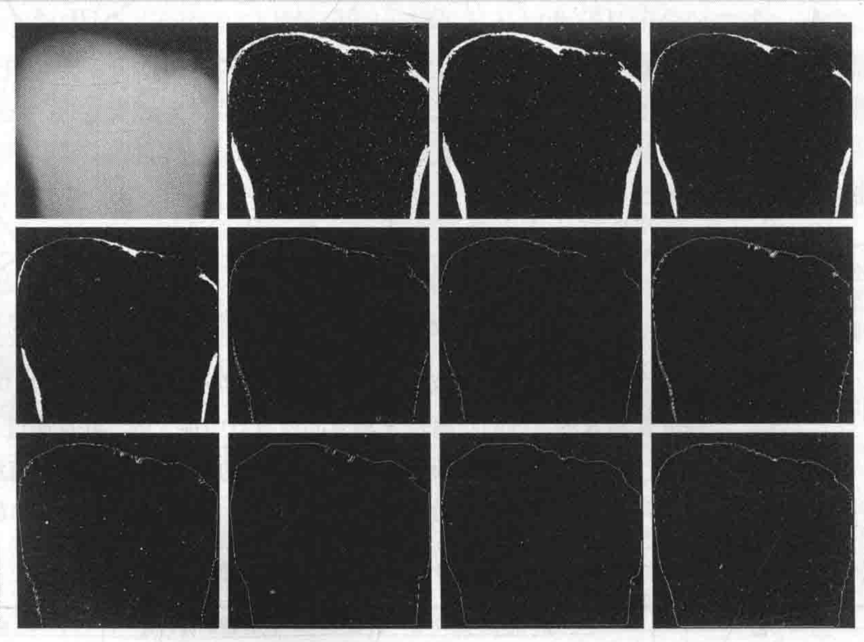
例 10.12 一幅图像的边界的多边形拟合。

图 10.30 显示了多边形拟合的一个更为实际的例子。图 10.30(a)中的输入图像是大小为 550×566 的人牙 X 射线图像，其灰度已标定到区间[0, 1]。该例的目的是提取牙齿的边界。图 10.30(b)是使用 Soble 模板和阈值 $T = 0.1$(最大灰度的 10%)得到的梯度图像。如预期的那样，X 射线图像中的噪声成分很高，因此，第一步是降噪。因为该图像是二值图像，故形态学技术可很好地适用于这一目的。图 10.30(c)显示了滤除大多数噪声后的结果，若在 3×3 邻域内有 5 个或更多个像素为 1，则置像素为 1，否则置像素为 0。虽然噪声已减少，但某些噪声点仍清晰可见。图 10.30(d)显示了形态学收缩后的结果，它进一步将噪声降低为孤立的点。这些噪声是用例 9.4 中描述的方式使用形态学滤波消除的[见图 10.30(e)]。在这一点上，图像由粗边界组成，这些边界可通过形态学骨架方法来加以细化，如图 10.30(f)所示。最后，图 10.30(g)显示了预处理中最后一步去除毛刺后的结果，如 9.5.8 节讨论的那样。

接着，我们用一个多边形来拟合图 10.30(g)中的点。图 10.30(h)~(j)显示了用多边形拟合算法的结果，其中阈值分别等于图像宽度的 0.5%, 1%和 2%(即 $T = 3, 6$ 和 12)。前两个结果较好地近似了边界，但第三个处在临界状态。所有这三种情况中，过度的锯齿清楚地表明需要进行边界平滑。图10.30(k)和(l)显示了一个一维均值模板分别与图(j)和图(h)中的边界卷积后的结果。所用的模板是元素为 1 的 1×31 阵列，约为

753 ~ 754

对应图像宽度的 5%。如所期望的那样，图 10.30(k)中的结果在保护重要形状特征方面仍是临界的(即右侧严重失真)。另一方面，图10.30(l)中的结果显示了有效的边界平滑，合理保留了形状特征。例如，图像以合理的保真度保留了左上方尖点的圆度和右上方尖点的细节。



a b c d
e f g h
i j k l

图 10.30 (a)一幅大小为 550×566 像素的人的牙齿图像；(b)梯度图像；(c)滤除大多数噪声的结果；(d)形态学收缩的结果；(e)形态学清除的结果；(f)骨架；(g)去除毛刺；(h)~(j)使用阈值约为图像宽度的 0.5%, 1%和 2%(即 $T=3, 6$ 和 12)的多边形拟合；(k)使用一个大小为 1×31(约为图像宽度的 5%)的一维均值滤波器平滑图(j)中边界后的结果；(l)使用相同滤波器平滑图(h)中边界后的结果

上例中的结果是使用本节讨论的多边形拟合算法所能得到的典型结果。该算法的优点是，实现简单，且产生的结果通常是可以接受的。11.1.3 节将讨论一种更完善的过程，即通过计算最小周长多边形可产生更接近的拟合。

使用霍夫变换的全局处理

前两节中讨论的方法可用于这样的场合，即属于各个目标的像素的知识部分可用。例如，在区域处理中，只有当我们知道它们是一个有意义区域边界的一部分时，连接给定的像素集才有意义。通常，我们不得不工作在无结构的环境中，这时我们拥有的只是一幅边缘图像，而没有感兴趣目标可能在哪里的知识。在这种情况下，所有像素都是连接的候选点，因此必须基于预先定义的全局性质接受或删除某些点。本节介绍一种基于像素集是否位于指定形状的曲线上的方法。一旦检测到，这些曲线就会形成边缘或感兴趣的区域边界。

给定一幅图像中的 n 个点，假设我们希望找到这些点中一个位于直线上的子集。一种可行的解决方法是，先找到所有由每对点确定的直线，然后寻找靠近特定直线的点的所有子集。这种方法涉及寻找 $n(n-1)/2 \sim n^2$ 条直线，然后对每个点，与所有直线执行$(n)(n(n-1))/2 \sim n^3$ 次比较。这种方法因为计算量太大而基本上没有应用价值。

754 ~ 755

霍夫[Hough 1962]提出一种替代方法，通常称为霍夫变换。考虑 xy 平面上的一点 (x_i, y_i) 和斜截式形式为 $y_i = ax_i + b$ 的一条直线。通过点 (x_i, y_i) 的直线有无数条，且对 a 和 b 的不同值，它们都满足方

程 $y_i = ax_i + b$ 。然而，将该等式写为 $b = -x_i a + y_i$ ，并考虑 ab 平面(也称为参数空间)，将得到固定点 (x_i, y_i) 的单条直线方程。此外，第二个点 (x_j, y_j) 在参数空间中也有一条与之相关联的直线，除非它们平行，否则这条直线会与和 (x_i, y_i) 相关联的直线相交于点(a', b')，其中 a'为斜率，b'是包含 xy 平面中点 (x_i, y_i) 和点 (x_j, y_j) 的直线的截距。事实上，这条直线上的所有点在参数空间中都有相交于点(a', b')的直线。图 10.31 说明了这些概念。

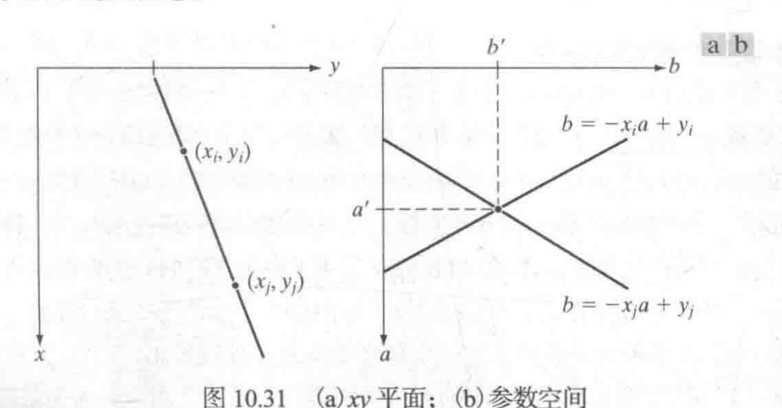


图 10.31 (a) xy 平面；(b) 参数空间

原理上，可以画出对应于 xy 平面中所有点 (x_k, y_k) 的参数空间直线，且空间中的主要直线可以在参数空间中通过确定的点来找到，参数空间的大量线在此点处相交。但这种方法的一个问题是，当该直线逼近垂直方向时，a(直线的斜率)会趋于无限大。解决该问题的方法之一是，使用一条直线的法线表示：

$$x\cos\theta + y\sin\theta = \rho \tag{10.2-38}$$

图 10.32(a)示例了参数 ρ 和 θ 的几何解释。水平直线有 $\theta = 0°$ ，ρ 等于正的 x 截距。类似地，垂直直线有 $\theta = 90°$ ，ρ 等于正的 y 截距；或者有 $\theta = -90°$ ，ρ 等于负的 y 截距。图 10.32(b)中的每条正弦曲线表示通过 xy 平面中某个特殊点 (x_k, y_k) 的一族直线。图 10.32(b)中的交点 (ρ', θ') 对应于图 10.32(a)中通过点 (x_i, y_i) 和点 (x_j, y_j) 的直线。

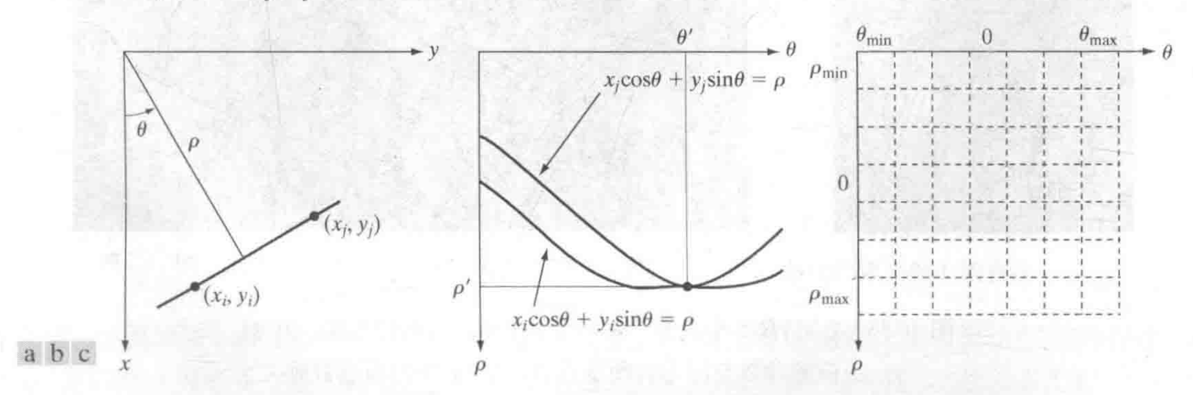


图 10.32 (a) xy 平面中直线的 (ρ, θ) 参数化；(b) $\rho\theta$ 平面中的正弦曲线，交点 (ρ', θ') 对应于通过 xy 平面中点 (x_i, y_i) 和点 (x_j, y_j) 的直线；(c) $\rho\theta$ 平面划分为累加单元

霍夫变换计算上的魅力在于可将 $\rho\theta$ 参数空间划分为所谓的累加单元，如图 10.32(c)所示，其中 (ρ_{min}, ρ_{max}) 和 $(\theta_{min}, \theta_{max})$ 是所期望的参数值范围：$-90° \leqslant \theta \leqslant 90°$ 和 $-D \leqslant \rho \leqslant D$ ，D 是图像中对角之间的最大距离。坐标(i, j)处的单元具有累加值 $A(i, j)$，它对应于与参数空间坐标 (ρ_i, θ_j) 相关联的正方形。最初，将这些单元置为零。然后，对于 xy 平面中的每个非背景点 (x_k, y_k) ，令 θ 等于

755 ~ 756

θ 轴上每个允许的细分值，同时使用方程 $\rho = x_k \cos\theta + y_k \sin\theta$ 解出对应的 ρ。对得到的 ρ 值进行四舍五入，得到沿 ρ 轴的最接近的允许单元值。若选择的一个 θ_p 值得到解 ρ_q，则令 $A(p,q) = A(p,q)+1$。在这一过程结束后，$A(i,j)$中的值 P 将意味着 xy 平面中有 P 个点位于直线 $x\cos\theta_j + y\sin\theta_j = \rho_i$ 上。$\rho\theta$ 平面中的细分数量决定了这些点的共线精度。可以证明（见习题 10.24），刚刚讨论的这种方法的计算次数与 xy 平面中非背景点的数量 n 呈线性关系。

例 10.13　基本霍夫变换性质的说明。

图 10.33 显示了基于式(10.2-38)的霍夫变换。图 10.33(a)显示了一幅大小为 101×101 像素并带有五个标记点的图像，图 10.33(b)显示了使用 ρ 轴和 θ 轴的单位细分将每个点映射到 $\rho\theta$ 平面上的结果。θ 的取值范围为 $\pm 90°$，ρ 轴的范围为 $\pm\sqrt{2}D$，其中 D 是图像中对角点间的距离。如图 10.33(c)所示，每条曲线都有不同的正弦曲线形状。点 1 的映射得到的水平直线，是具有零幅值的正弦曲线的一种特殊情形。

图 10.33(b)中标为 A（不要与累加值混淆）和 B 的点显示了霍夫变换的共线性检测性质。点 A 表示对应于 xy 图像平面内点 1, 3 和 5 的曲线的交点。点 A 的位置指出这三个点位于一条过原点（$\rho = 0$）且方向为 45° 的直线上［见图 10.32(a)］。类似地，在参数空间中相交于点 B 的曲线指出，点 2, 3 和 4 位于方向为 −45° 且与原点的距离为 $\rho = 71$（即从图像原点到对角的对角线距离的一半，已四舍五入为最接近的整数值）的直线上。最后，图 10.33(b)中标为 Q, R 和 S 的点说明了这样一个事实，即霍夫变换展示了在参数空间左边缘和右边缘处的一种反射邻接关系。这一性质是 θ 和 ρ 在 $\pm 90°$ 边界改变符号的结果。

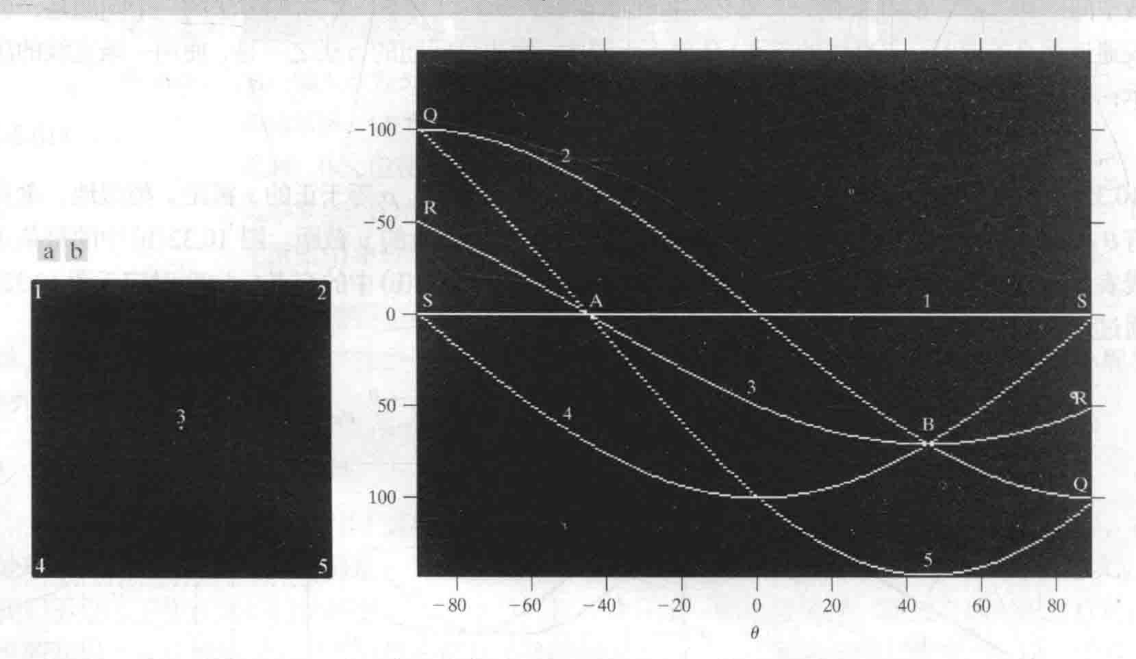


图 10.33　(a)包含 5 个点的、大小为 101×101 像素的图像；(b)对应的参数空间［为便于观看，图(a)中的点已被放大］

尽管到目前为止我们关注的重点始终在直线上，但霍夫变换也适用于形如 $g(\boldsymbol{v},\boldsymbol{c}) = 0$ 的任何函数，其中 $\boldsymbol{v}$ 是坐标向量，$\boldsymbol{c}$ 是系数向量。例如，位于圆

$$(x-c_1)^2 + (y-c_2)^2 = c_3^2 \tag{10.2-39}$$

上的点可以使用刚才讨论的方法来检测。不同之处是存在 3 个参数（c_1, c_2 和 c_3），在三维参数空间中，这三个参数导致了类似立体的单元和形如 $A(i,j,k)$的累加器。过程是，增加 c_1 和 c_2，求出满足式(10.2-39)的 c_3，并更新与三元组(c_1, c_2, c_3)相关联的累加单元。很明显，霍夫变换的复杂性取决

756 ~ 758

于给定函数表达式中的坐标和系数数量。霍夫变换有可能进一步推广到检测无简单解析表达式的曲线，如变换为灰度级图像那样的应用。本章末尾包含了处理这些扩展功能的一些参考文献。

现在回到边缘连接问题。基于霍夫变换的一种连接方法如下：

1. 使用本节中讨论的任何技术得到一幅二值图像。
2. 指定$\rho\theta$平面中的细分。
3. 对像素高度集中的地方检验其累加单元的数量。
4. 检验选中单元中像素间的关系（主要针对连续性）。

在这种情况下，连续性通常以对应于给定的累加单元计算不连续像素间的距离为基础。若缝隙的长度比指定的阈值小，则连接与给定单元相关联的一条直线中的缝隙。注意，仅以方向为基础聚合直线的事实是一个可用于整个图像的全局概念，它仅要求我们考察与指定累加单元相关联的像素。与前两节中讨论的方法相比，这是一个明显的优点。下例说明了这些概念。

例 10.14　使用霍夫变换连接边缘。

图 10.34(a) 显示了一幅航拍的机场图像。该例的目的是使用霍夫变换提取主要跑道的两条边。解决这样的问题可能很重要，例如，涉及飞行器自动导航的应用。

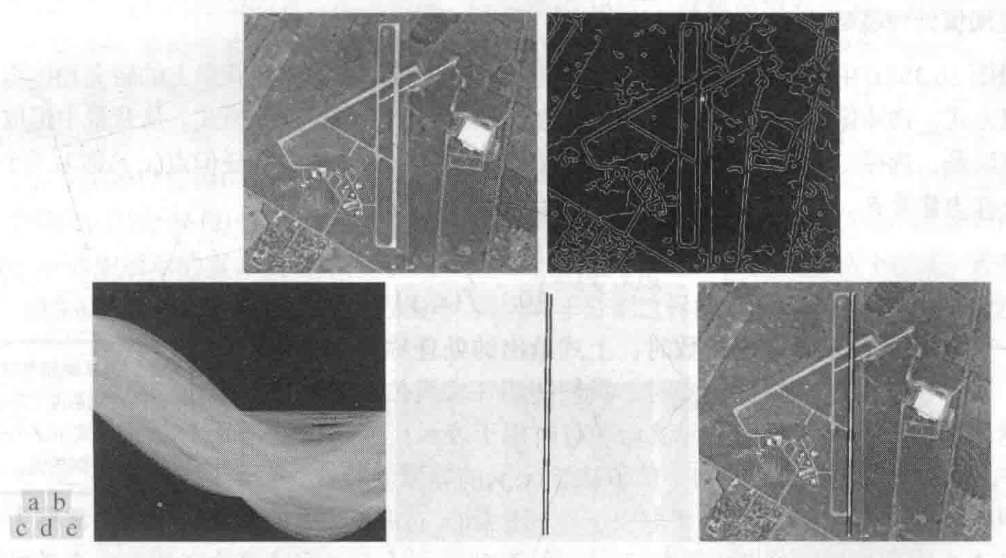


图 10.34　(a) 一幅大小为 502×564 的机场航拍图像；(b) 使用坎尼算法得到的边缘图像；(c) 霍夫参数空间（方框强调了与长垂直线相关联的点）；(d) 图像平面中对应于方框所强调的点的直线；(e) 已叠加到原图像上的直线

第一步是得到一幅边缘图像。图10.34(b) 显示了使用坎尼算法得到的边缘图像，算法中使用的参数和过程与例10.9 中的相同。为计算霍夫变换，使用 10.2.5 节或 10.2.6 节中讨论的任何边缘检测技术，可得到类似的结果。图 10.34(c) 显示了θ以 1°递增和ρ以 1 像素递增得到的霍夫参数空间。

跑道偏离正北方向约 1°，所以我们选取对应于±90° 并包含最高数量的单元，因为跑道在这些方向面向最长的线。图 10.34(c) 的边缘上的小白框强调了这些单元。结合图 10.33(b)，如先前提及的那样，霍夫变换展示了边缘处的邻接性。解释这一性质的另一种方法是，面向+90° 的一条线和面向 −90° 的一条线是等价的（即它们是两条垂直线）。图10.34(d) 显示了对应于刚才讨论的两个累加单元的直线，图 10.34(e) 显示了已叠加到原图像上的这些直线。这些直线是通过连接不超过图像高度的 20%（约 100 个像素）的所有间隙得到的。这些直线清楚地对应于感兴趣跑道的边缘。

758 ~ 759

注意，解决这一问题的唯一关键知识是，跑道的方向和观察者相对于跑道的位置。换句话说，自动导航飞行器应该知道，如果跑道朝北，且飞行器的飞行方向也是北，跑道会垂直地出现在图像中。其他相关的方向可采用类似的方式处理。全世界的跑道的方向在飞行图上都是可用的，且使用GPS(全球定位系统)可很容易得到飞行方向的信息。该信息还可用于计算飞行器和跑道之间的距离，这样就允许我们估计这些参数，例如相对于图像尺寸的直线的预期长度，就像我们在该例中所做的那样。

10.3 阈值处理

由于阈值处理直观、实现简单且计算速度块，因此图像阈值处理在图像分割应用中处于核心地位。3.1.1节已介绍了阈值处理，自那时起在各种讨论中我们就使用过阈值处理。本节将更为正式地介绍阈值处理。

10.3.1 基础知识

在前面的章节中，我们采用首先寻找边缘线段，然后将这些线段连接为边界的方法来识别区域。本节讨论基于灰度值和/或灰度值的特性来将图像直接划分为区域的技术。

灰度阈值处理基础

假设图10.35(a)中的灰度直方图对应于图像$f(x, y)$，该图像由暗色背景上的较亮物体组成，以这样的组成方式，物体像素和背景像素所具有的灰度值组合成了两种支配模式。从背景中提取物体的一种明显方法是，选择一个将这些模式分开的阈值T。然后，$f(x, y) > T$的任何点(x, y)称为一个*对象点*；否则该点称为*背景点*。换句话说，分割后的图像$g(x, y)$由下式给出：

$$g(x, y)=\begin{cases}1, & f(x, y)>T \\ 0, & f(x, y) \leqslant T\end{cases} \tag{10.3-1}$$

> 尽管我们遵从使用灰度0来表示背景像素而使用灰度1来表示物体像素的约定，但在式(10.3-1)中也可使用任何两个明显不同的值。

当T是一个适用于整个图像的常数时，上式给出的处理称为*全局阈值处理*。当T值在一幅图像上改变时，我们使用*可变阈值处理*这一术语。术语*局部阈值处理*或*区域阈值处理*有时用于表示可变阈值处理，此时，图像中任何点(x, y)处的T值取决于(x, y)的邻域的特性(例如邻域中的像素的平均灰度)。若T取决于空间坐标(x, y)本身，则可变阈值处理通常称为*动态阈值处理*或*自适应阈值处理*。这些术语的应用并不普遍，在关于图像处理的文献中，读者可能会看到它们交替地使用。

图10.35(b)显示了一个更为困难的阈值处理问题，它包含有三个支配模式的直方图，例如，这三个支配模式对应于暗色背景上的两个明亮物体。这里，若$f(x, y) \leqslant T_1$，则多阈值处理把点(x, y)分类为背景；若$T_1 < f(x, y) \leqslant T_2$，则分类为一个物体；若$f(x, y) > T_2$，则分类为另一个物体。即分割的图像由下式给出：

$$g(x, y)=\begin{cases}a, & f(x, y)>T_2 \\ b, & T_1<f(x, y) \leqslant T_2 \\ c, & f(x, y) \leqslant T_1\end{cases} \tag{10.3-2}$$

式中，a, b和c是任意三个不同的灰度值。在10.3.6节中，我们将讨论双阈值处理。要求两个以上阈值的分割问题很难解决(通常是不可能的)，而较好的结果通常可用其他方法得到，如10.3.7节中讨论的可变阈值处理方法或10.4节中讨论的区域生长方法。

759 ~ 761

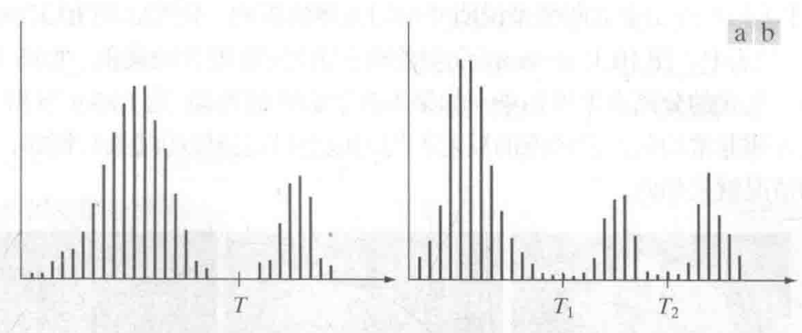


图 10.35　可被(a)单阈值和(b)双阈值分隔的灰度直方图

基于前面的讨论，我们可以凭直觉推断灰度阈值的成功与否直接关系到可区分直方图模式的波谷的宽度和深度。而影响波谷特性的关键因素是：(1)波峰间的间隔(波峰离得越远，分离这些模式的机会越好)；(2)图像中的噪声内容(模式随噪声的增加而展宽)；(3)物体和背景的相对尺寸；(4)光源的均匀性；(5)图像反射特性的均匀性。

图像阈值处理中噪声的作用

作为噪声如何影响一幅图像的直方图的说明，考虑图10.36(a)。这幅简单的合成图像中没有噪声，因此，其直方图由两个波峰模式组成，如图 10.36(d)所示。将该图像分割为两个区域很简单，只需将一个阈值放到两个模式之间的任何位置。图10.36(b)显示了被均值为零、标准差为 10 个灰度级的高斯噪声污染了的原始图像。尽管相应的直方图模式现在较宽［见图10.36(e)］，但它们的间隔足够大，故它们之间波谷的深度足以使得两个模式更易于分开。放在两个波峰之间的中间位置的一个阈值就可以很好地分割该图像。图10.36(c)显示了受到均值为零、标准差为 50 个灰度级的高斯噪声污染该图像的结果。如图10.36(f)中所示的直方图那样，现在情况是如此严重，以至于无法区分两个模式。若没有附加的处理(如 10.3.4 节和 10.3.5 节中讨论的方法)，我们对于分割这样的图像寻找合适的阈值就没有什么希望。

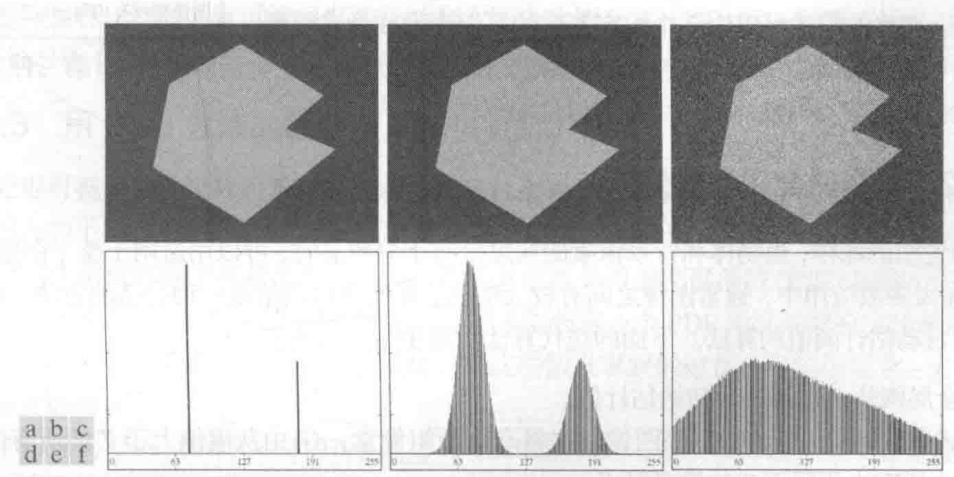


图 10.36　(a)无噪声的 8 比特图像；(b)带有均值为零、标准差为 10 个灰度级的加性高斯噪声的图像；(c)带有均值为零、标准差为 50 个灰度级的加性高斯噪声的图像；(d)~(f)相应的直方图

光照和反射的作用

图 10.37 说明了光照对一幅图像的直方图的影响。图 10.37(a)是来自图 10.36(b)的噪声图像，图 10.37(d)显示了它的直方图。像先前那样，使用一个单阈值很容易分割该图像。我们可以通过一个

761~762

可变的灰度函数乘以图 10.37(a)中的图像来说明非均匀光照的影响，例如，图 10.37(b)的斜坡灰度，其直方图显示在图 10.37(e)中。图 10.37(c) 显示了这幅图像与该阴影模式的乘积。如图 10.37(f) 所示，波峰之间的较深波谷在模式的分离点处被污染，如果不进行额外的处理(见 10.3.4 节和 10.3.5 节)，分离是不可能的。如果光照非常均匀，但图像的反射不均匀也会得到类似的结果。例如，物体的表面或背景自然反射变化的情况就是如此。

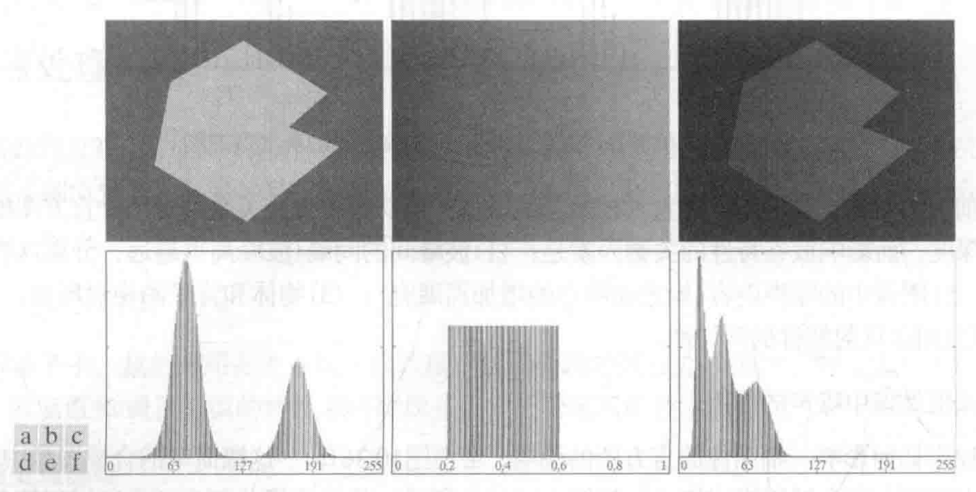


图 10.37 (a)带有噪声的图像；(b)在[0.2, 0.6]范围内的灰度斜坡图像；(c)图(a)和图(b)的乘积；(d)~(f)相应的直方图

前面章节中的关键是，在使用阈值或其他分割技术的图像成功分割中，照明和反射起到了核心作用。因此，在可能的情况下，在解决分割的问题中，首先应考虑如何控制这些参数。当不可能控制这些参数时，有三种针对该问题的基本解决方法。第一种方法是直接校正这种阴影模式。例如，非均匀(但固定的)光照可以用相反的模式与图像相乘来校正，相反的模式可以用一个恒定灰度的平坦表面成像来得到。第二种方法可通过用诸如 9.6.3 节中介绍的顶帽变换处理来尝试校正全局阴影模式。第三种方法是使用可变阈值近似处理非均匀性，如 10.3.7 节讨论的那样。

> 理论上，斜坡图像的直方图是均匀的。实际上，实现完美的均匀性取决于图像的尺寸和灰度比特数。例如，一幅大小为 256×256 的 256 级斜坡图像有一个均匀的直方图，但大小为 256×257 的具有相同灰度级数的斜坡图像则没有均匀的直方图。

10.3.2 基本的全局阈值处理

如前节说明的那样，当物体和背景像素的灰度分布十分明显时，可以用适用于整个图像的单个(全局)阈值。在大多数应用中，通常图像之间有较大变化，即使全局阈值是一种合适的方法，也需要有能对每幅图像自动估计阈值的算法。下面的迭代算法可用于这一目的：

1. 为全局阈值 T 选择一个初始估计值。
2. 在式(10.3-1)中用 T 分割该图像。这将产生两组像素：G_1 由灰度值大于 T 的所有像素组成，G_2 由所有小于等于 T 的像素组成。
3. 对 G_1 和 G_2 的像素分别计算平均灰度值(均值) m_1 和 m_2。
4. 计算一个新的阈值：

$$T = \frac{1}{2}(m_1 + m_2)$$

5. 重复步骤 2 到步骤 4，直到连续迭代中的 T 值间的差小于一个预定义的参数 ΔT 为止。

762~764

当与物体和背景相关的直方图模式间存在一个相当清晰的波谷时，这个简单的算法工作得很好。在速度是一个重要因素的情形下，参数 ΔT 用于控制迭代的次数。通常，ΔT 越大，则算法执行的迭代次数越少。所选的初始阈值必须大于图像中的最小灰度级而小于最大灰度级（见习题 10.28）。图像的平均灰度对于 T 来说是较好的初始选择。

例 10.15　全局阈值处理

图 10.38 显示了用前述算法以阈值估计为基础分割的例子。图 10.38(a) 是原图像，图 10.38(b) 是该图像的直方图，该直方图显示有一个明显的波谷。用 $T = m$（平均图像灰度）开始，并令 $\Delta T = 0$，应用前述迭代算法经过 3 次迭代后，得到阈值 $T = 125.4$。图 10.38(c) 显示了使用阈值 $T = 125$ 来分割原图像得到的结果。如期望的那样，从直方图清晰的模式分离中，物体和背景间的分割相当有效。

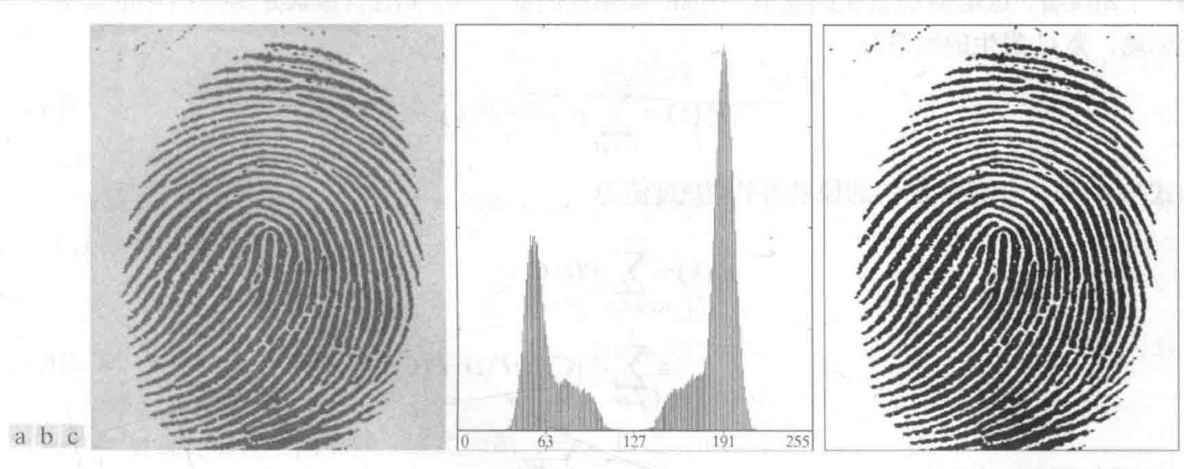


图 10.38　(a) 带噪声的指纹；(b) 直方图；(c) 使用全局阈值对图像分割的结果（为清晰起见，边界是加上去的）（原图像由美国国家标准和技术研究所提供）

通过对输入图像成功地进行阈值处理并在每一步计算均值说明了前述算法，因为以这种方式来介绍该算法更为直观。然而，有可能根据图像的直方图表达的全部计算开发一种更有效的过程，它仅仅计算一次（见习题 10.26）。

10.3.3　用 Otsu 方法的最佳全局阈值处理

阈值处理可视为一种统计决策理论问题，其目的是在把像素分配给两个或多个组（也称为分类）的过程中使引入的平均误差最小。这一问题已知有一个精致的闭合形式的解，称为贝叶斯决策规则（见 12.2.2 节）。这个解仅以两个参数为基础：每一类的灰度级的概率密度函数（PDF）和给定应用中每一类出现的概率。遗憾的是，估计 PDF 并不容易，通常采取假定一个 PDF 形式的可行方法来简化该问题，例如，假设它们是高斯函数。即便采用了简化，用这些假设求解的过程也可能很复杂，并且对实际应用也不总是合适的。

本节中讨论的 Otsu 方法（Otsu[1979]）是另一种有吸引力的方案。该方法在类间方差最大的情况下是最佳的，这是统计鉴别分析中所用的测度。基本概念是，好阈值分类就其像素灰度值而论，应是截然不同的，反过来说，就其灰度值而言给出最好的类间分离的阈值就是最好的（最佳的）阈值。除了其最佳性之外，Otsu 方法还有一个重要的特性，即它完全以在一幅图像的直方图上执行计算为基础，直方图是很容易得到的一维阵列。

令 $\{0, 1, 2, \cdots, L-1\}$ 表示一幅大小为 $M \times N$ 像素的数字图像中的 L 个不同的灰度级，n_i 表示灰度级为 i 的像素数。图像中的像素总数为 $MN = n_0 + n_1 + n_2 + \cdots + n_{L-1}$。归一化的直方图（见 3.3 节）具有分

764 ~ 765

量 $p_i = n_i / MN$ ，由此有

$$\sum_{i=0}^{L-1} p_i = 1, \qquad p_i \geqslant 0 \tag{10.3-3}$$

现在，假设我们选择一个阈值 $T(k) = k, 0 < k < L-1$，并使用它把输入图像阈值化处理为两类 C_1 和 C_2 ，其中，C_1 由图像中灰度值在区间[0, k]内的所有像素组成，C_2 由灰度值在区间[k+1, L−1]内的所有像素组成。用该阈值，像素被分到类 C_1 中的概率 $P_1(k)$由如下的累积和给出：

$$P_1(k) = \sum_{i=0}^{k} p_i \tag{10.3-4}$$

换一个角度看，这是类 C_1 发生的概率。例如，如果我们置 $k=0$，则任何像素分到类 C_1 中的概率为零。类似地，类 C_2 发生的概率为

$$P_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - P_1(k) \tag{10.3-5}$$

由式(3.3-18)，分配到类 C_1 的像素的平均灰度值为

$$\begin{aligned} m_1(k) &= \sum_{i=0}^{k} iP(i/C_1) \\ &= \sum_{i=0}^{k} iP(C_1/i)P(i)/P(C_1) \\ &= \frac{1}{P_1(k)} \sum_{i=0}^{k} ip_i \end{aligned} \tag{10.3-6}$$

式中，$P_1(k)$由式(10.3-4)给出。式(10.3-6)第一行中的 $P(i/C_1)$ 项是值 i 的概率，已知 i 来自类 C_1。该式中的第二行来自贝叶斯公式：

$$P(A/B) = P(B/A)P(A)/P(B)$$

第三行遵循这样一个事实，即给定 i 的 C_1 的概率 $P(C_1/i)$ 为 1，因为我们只处理来自类 C_1 的 i 值。此外，$P(i)$ 是第 i 个值的概率，很简单，它是直方图 $p(i)$ 的第 i 个分量。最后，$P(C_1)$ 是类 C_1 的概率，由式(10.3-4)我们知道它等于 $P_1(k)$ 。

类似地，分配到类 C_2 中的像素的平均灰度值为

$$\begin{aligned} m_2(k) &= \sum_{i=k+1}^{L-1} iP(i/C_2) \\ &= \frac{1}{P_2(k)} \sum_{i=k+1}^{L-1} ip_i \end{aligned} \tag{10.3-7}$$

级 k 的累加均值由下式给出：

$$m(k) = \sum_{i=0}^{k} ip_i \tag{10.3-8}$$

而整个图像的平均灰度(即全局均值)由下式给出：

765
~
766

$$m_G = \sum_{i=0}^{L-1} ip_i \tag{10.3-9}$$

下面两个公式的正确性可通过直接代入前面的结果来验证：

$$P_1 m_1 + P_2 m_2 = m_G \tag{10.3-10}$$

和

$$P_1 + P_2 = 1 \tag{10.3-11}$$

其中，为有利于清楚地说明，我们暂时忽略了 k 项。

为了评价级别 k 处的阈值的“质量”，我们使用归一化的无量纲矩阵：

$$\eta = \frac{\sigma_B^2}{\sigma_G^2} \tag{10.3-12}$$

式中，σ_G^2 是全局方差［即图像中所有像素的灰度方差，如式(3.3-19)给出的那样］，

$$\sigma_G^2 = \sum_{i=0}^{L-1} (i - mG)^2 p_i \tag{10.3-13}$$

σ_B^2 为类间方差，它定义为

$$\sigma_B^2 = P_1 (m_1 - m_G)^2 + P_2 (m_2 - m_G)^2 \tag{10.3-14}$$

该表达式还可写为

$$\begin{aligned} \sigma_B^2 &= P_1 P_2 (m_1 - m_2)^2 \\ &= \frac{(m_G P_1 - m)^2}{P_1(1 - P_1)} \end{aligned} \tag{10.3-15}$$

式中，m_G 和 m 如先前说明的那样。该式的第一行来自式(10.3-14)、式(10.3-10)和式(10.3-11)。第二行来自式(10.3-5)到式(10.3-9)。这种形式在计算上更有效一些，因为全局均值 m_G 仅计算一次，故对 k 的任何值仅需要计算两个参数 m 和 P_1。

> 式(10.3-15)的第二步仅在 P_1 大于 0 且小于 1 时才有意义，鉴于式(10.3-11)，这表明 P_2 必须满足相同的条件。

从式(10.3-15)的第一行我们可以看出，两个均值 m_1 和 m_2 彼此隔得越远，σ_B^2 越大，这表明类间方差是类间的可分性测度。因为 σ_G^2 是一个常数，由此得出 η 也是一个可分性测度，且最大化这一测度等价于最大化 σ_B^2。然后，目标是确定阈值 k，它最大化类间方差，就像本节开始时说明的那样。注意，式(10.3-12)隐含假设了 $\sigma_G^2 > 0$。仅当图像中的所有灰度级相同时，这一方差才为零，这意味着仅存在一类像素。同样，这也意味着对于常数图像有 $\eta = 0$，因为来自其自身单个类的可分性为零。

再次引入 k，我们有最终结果：

$$\eta(k) = \frac{\sigma_B^2(k)}{\sigma_G^2} \tag{10.3-16}$$

和

$$\sigma_B^2(k) = \frac{[m_G P_1(k) - m(k)]^2}{P_1(k)[1 - P_1(k)]} \tag{10.3-17}$$

从而最佳阈值是 k^*，它最大化 $\sigma_B^2(k)$：

$$\sigma_B^2(k^*) = \max_{0 \le k \le L-1} \sigma_B^2(k) \tag{10.3-18}$$

换句话说，为寻找 k^*，我们对 k 的所有整数值［保持上述 $0 < P_1(k) < 1$ 的条件］简单地计算式(10.3-18)，并选取使得 $\sigma_B^2(k)$ 最大的 k 值。如果 $\sigma_B^2(k)$ 的最大值对应于多个 k 值，习惯上是对于使 $\sigma_B^2(k)$ 最大的

767~768

k 的各个值做平均。可以证明(见习题 10.33)，$0 < P_1(k) < 1$ 时，总存在一个最大值。对于所有的 k 值，计算式(10.3-17)和式(10.3-18)的代价相对较小，因为 k 所具有的整数值的最大数量是 L。

一旦得到 k^*，输入图像就可像之前那样进行分割：

$$g(x, y) = \begin{cases} 1, & f(x, y) > k^* \\ 0, & f(x, y) \leqslant k^* \end{cases} \tag{10.3-19}$$

式中，$x = 0, 1, 2, \cdots, M-1$ 和 $y = 0, 1, 2, \cdots, N-1$。注意，仅用 $f(x, y)$ 的直方图就可以得到计算式(10.3-17)所需的所有参量。除了最佳阈值外，与分割图像有关的其他信息可从直方图中提取。例如，在最佳阈值处计算出的类的概率 $P_1(k^*)$ 和 $P_2(k^*)$，指出了经阈值处理后的图像中，由该类(像素组)所占据的面积部分。类似地，均值 $m_1(k^*)$ 和 $m_2(k^*)$ 是原图像中类的平均灰度的估计。

在最佳阈值处计算的归一化测度 η，即 $\eta(k^*)$，可用于得到类别可分性的定量估计，同时给出了很容易对一幅给定图像进行阈值处理的概念。这一测度的值域为

> 尽管我们的兴趣是最佳阈值 k^* 处的 η 值，但该不等式对于范围[0, $L-1$]内的任何 k 值均成立。

$$0 \leqslant \eta(k^*) \leqslant 1 \tag{10.3-20}$$

如先前提到的那样，仅由单一且恒定灰度级的图像就能得到下界，且仅由灰度等于 0 和 $L-1$ 的二值图像就能得到上界(见习题 10.34)。

Otsu 算法小结如下：

1. 计算输入图像的归一化直方图。使用 $p_i, i = 0, 1, 2, \cdots, L-1$ 表示该直方图的各个分量。
2. 用式(10.3-4)，对于 $k = 0, 1, 2, \cdots, L-1$，计算累积和 $P_1(k)$。
3. 用式(10.3-8)，对于 $k = 0, 1, 2, \cdots, L-1$，计算累积均值 $m(k)$。
4. 用式(10.3-9)计算全局灰度均值 m_G。
5. 用式(10.3-17)，对于 $k = 0, 1, 2, \cdots, L-1$，计算类间方差 $\sigma_B^2(k)$。
6. 得到 Otsu 阈值 k^*，即使得 $\sigma_B^2(k)$ 最大的 k 值。如果最大值不唯一，用相应检测到的各个最大值 k 的平均得到 k^*。
7. 在 $k = k^*$ 处计算式(10.3-16)，得到可分性测度 η^*。

下面的例子说明了前述概念。

例 10.16　使用 Otsu 方法的最佳全局阈值处理。

图 10.39(a)显示了聚合细胞的光学显微图像，图 10.39(b)显示了其直方图。该例子的目的是从背景中分割出分子。图 10.39(c)是使用前节给出的基本全局阈值处理算法得到的结果。因为直方图没有明显的波谷，且背景和物体间的灰度差别很小，所以算法未完成期望的分割。图 10.39(d)显示了使用 Otsu 方法得到的结果。该结果明显好于图 10.39(c)中的结果。使用基本算法计算出的阈值是 169，而使用 Otsu 方法计算出的阈值是 181，后者更接近图像中定义为细胞的较亮区域。可分性测度 η 是 0.467。

> 聚合细胞是使用聚合物人工生成的细胞。人眼看不到聚合细胞，这种细胞可用于向人体内的目标区域输送药物。

有趣的是，将 Otsu 方法应用于例 10.15 中的指纹图像，得到的阈值和可分性测度分别为 125 和 0.944。该阈值与使用基本算法得到的阈值(已四舍五入为最接近的整数)大小相同。对给定该直方图的特性，这一结果并不意外。事实上，由于原来的模式间的分离度相对较大，并且它们之间的波谷较深，所以可分性测度较高。

768 ~ 769

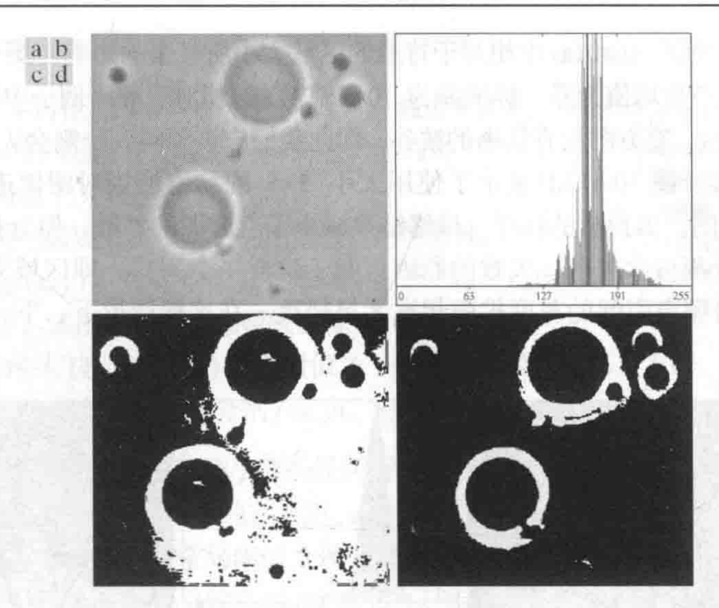


图 10.39　(a)原图像；(b)直方图(为突出低值中的细节，高峰已被裁剪)；(c)用 10.3.2 节的基本全局算法得到的分割结果；(d)用 Otsu 方法得到的结果(原图像由宾夕法尼亚大学的 Daniel A. Hammer 教授提供)

10.3.4　用图像平滑改善全局阈值处理

如图 10.36 中解释的那样，噪声会将简单的阈值处理问题变为不可解决的问题。当噪声不能在源头减少，并且阈值处理又是所选择的分割方法时，那么通常能增强性能的一种技术是，在阈值处理之前平滑图像。我们用一个例子来说明这一方法。

图 10.40(a)是来自图 10.36(c)的图像，图 10.40(b)显示了其直方图，图 10.40(c)是使用 Otsu 方法对图像进行阈值处理后的结果。白色区域中的每个黑点和黑色区域中的每个白点是阈值处理的误差，所以分割很不成功。图 10.40(d)显示了使用一个大小为 5×5(图像大小为 651×814 像素)的均值模板平滑噪声图像后的结果，图 10.40(e)是其直方图。由于平滑处理，直方图形状的改进很明显，因此我们预料对平滑后的图像进行阈值处理所得到的结果将近于完美。如图 10.40(f)所示，事实的确如此。经平滑和分割后的图像中，物体和背景间的边界稍微有点失真是由于对边界的模糊造成的。事实上，我们应该预料到，对一幅图像平滑多越多，分割后的结果中的边界误差就越大。

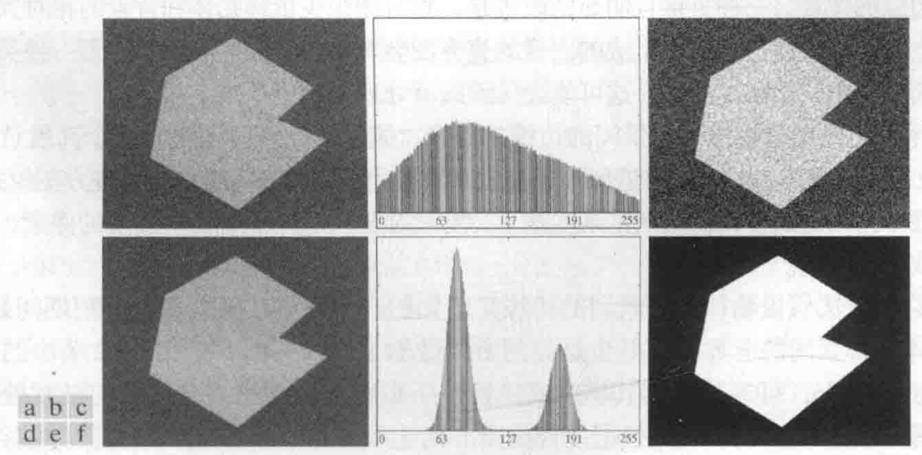


图 10.40　(a)来自图 10.36 的噪声图像；(b)该图像的直方图；(c)用 Otsu 方法得到的结果；(d)用一个大小为 5×5 的均值模板平滑噪声图像的结果；(e)结果图像的直方图；(f)用 Otsu 方法阈值处理后的结果

769 ~ 770

下面我们考虑降低图 10.40(a)中相对于背景的区域大小所带来的影响。图 10.41(a)显示了该结果。这幅图像中的噪声是均值为零、标准差为 10 个灰度级(相对于前一例子中的 50)的加性高斯噪声。如图10.41(b)所示，直方图没有清晰的波谷，因此我们应该预料到分割会失败，图10.41(c)中的结果确认了这一事实。图 10.41(d)显示了使用大小 5×5 的均值模板对图像进行平滑后的结果，图 10.41(e)是其直方图。如预料的那样，最终结果减小了直方图的扩散，但分布仍呈单峰形式。如图 10.41(f)所示，分割再次失败。失败的原因可归于这样一个事实，即区域太小，以至于该区域对直方图的贡献与由噪声引起的灰度扩散相比无足轻重。在这种情形下，下一节讨论的方法可能更成功。

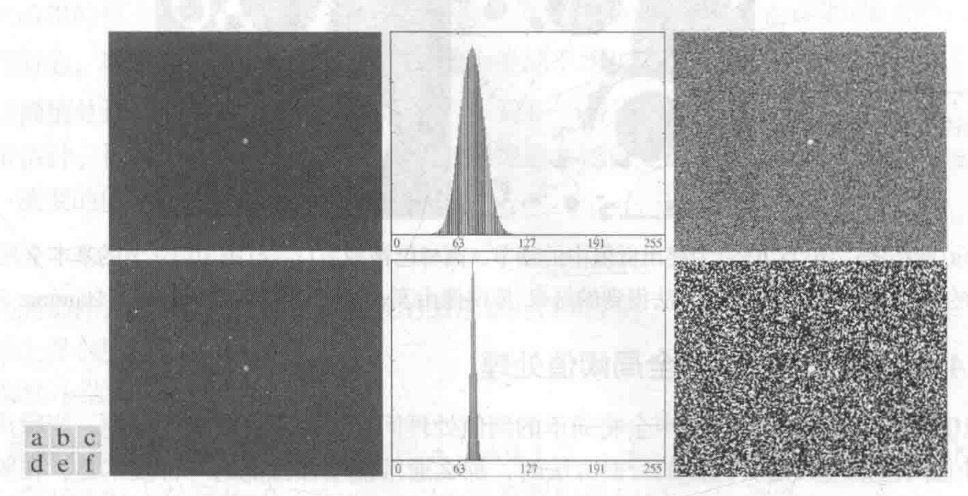


图 10.41 (a)噪声图像；(b)其直方图；(c)用 Otsu 方法得到的结果；(d)用大小为 5×5 的均值模板平滑噪声图像后的结果；(e)结果图像的直方图；(f)用 Otsu 方法对图像进行阈值处理后的结果。两种情形下，阈值处理均失败

10.3.5 利用边缘改进全局阈值处理

基于前四节的讨论，我们得出如下结论：如果直方图的波峰是高、窄、对称的，且被深波谷分开，则选取一个“较好”阈值的机会相当大。改进直方图形状的一种方法是，仅考虑那些位于或靠近物体和背景间的边缘的像素。一种直接且明显的改进是，直方图很少依赖物体和背景的相对大小。例如，大背景区域上一个小物体(或相反)组成的图像的直方图会有一个较大的波峰，因为一种类型的像素高度集中。在前一节中我们已了解到，这可能会导致阈值处理失败。

如果仅用位于或接近物体和背景间的边缘的像素，则得到的直方图将有几个高度近似相同的波峰。此外，像素位于物体上的概率将近似等于其位于背景中的概率，从而改进了直方图模式的对称性。最后，如下段中指出的那样，使用满足某些基于梯度和拉普拉斯算子的简单测度的像素，有加深直方图波峰间的波谷的倾向。

刚才讨论的方法假设物体和背景间的边缘是已知的。很明显，该信息在分割期间是不可用的，因为寻找物体和背景间的分界线正好也是分割所要做的工作。然而，根据 10.2 节中的讨论，一个像素是否位于边缘上，可通过计算其梯度或拉普拉斯来确定。例如，在边缘的过渡处，拉普拉斯的平均值为零(见图 10.10)，因此，根据拉普拉斯准则选取的像素所形成的直方图的波谷应该很少。这一特性往往会产生上面讨论的较深波谷。实际上，将用梯度图像得到的结果与使用拉普拉斯图像得到的结果相比较，后者比较有利，因为它在计算上更有吸引力，同时也是一个各向同性边缘检测器。

770
~
772

前述讨论可小结为如下算法，其中$f(x, y)$是输入图像：

1. 采用 10.2 节讨论的任何一种方法来计算一幅边缘图像，无论是$f(x, y)$梯度的幅度还是拉普拉斯的绝对值均可。
2. 指定一个阈值 T。
3. 用步骤 2 中的阈值对步骤 1 中的图像进行阈值处理，产生一幅二值图像 $g_T(x, y)$。在从$f(x, y)$中选取对应于“强”边缘像素的下一步中，该图像用做一幅模板图像。
4. 仅用$f(x, y)$中对应于 $g_T(x, y)$ 中像素值为 1 的位置的像素计算直方图。
5. 用步骤 4 中的直方图全局地分割$f(x, y)$，例如使用 Otsu 方法。

有可能修改该算法以便同时使用梯度图像的幅度和拉普拉斯图像的绝对值。在这种情况下，我们会为每幅图像指定一个阈值，并求两个结果的逻辑“或”(OR)，以得到标记图像。当需要对有效的边缘点施加更多的控制时，这种方法很有用。

若将 T 设置为小于边缘图像的最小值的任何值，根据式(10.3-1)，$g_T(x, y)$ 将由全 1 组成，这意味着$f(x, y)$的所有像素将用于计算图像的直方图。在这种情况下，前述算法变成使用原始图像的直方图的全局阈值处理。习惯上以百分比的形式来指定 T 值，百分比通常设置得较高(如 90%)，以便在计算中用梯度/拉普拉斯图像中的较少像素。下面的例子说明了刚才讨论的概念。第一个例子使用梯度，第二个例子使用拉普拉斯。在两个例子中，无论使用哪种方法都可得到类似的结果。重要的问题是生成一幅合适的导数图像。

第 n 个百分比是大于给定集合内所有数字的 n%的最小数字。例如，如果读者在一次测验中得到了 95 分，且该分数大于参加此次测验的所有学生中的 85%的学生得到的分数，则读者在测试分数中位于第 85 个百分比处。

例 10.17　用以梯度为基础的边缘信息改进全局阈值处理。

图 10.42(a)和(b)显示了来自图 10.41 的图像和直方图。我们可以看到，这幅图像不能先进行平滑后进行阈值处理的方式来分割。这个例子的目是解决利用边缘信息的问题。图 10.42(c)是在 99.7%处经阈值处理后的梯度幅度图像。图 10.42(d)是用输入图像与这个(模板)图像相乘形成的图像。图 10.42(e)是图 10.42(d)中非零元素的直方图。注意，该直方图具有先前讨论的重要特性，即它有被一个较深波谷分开的对称模式。这样，尽管原始带噪图像的直方图未提供成功进行阈值处理的机会，但图 10.42(e)中的直方图表明，从背景中提取出小物体的阈值处理的确是可能的。图 10.42(f)中的结果表明事实的确如此。这幅图像是这样得到的：以图 10.42(e)的直方图为基础，用 Otsu 方法得到一个阈值，然后将该阈值全局地应用到图 10.42(a)中的带噪图像。结果近于完美。

例 10.18　用拉普拉斯为基础的边缘信息改进全局阈值处理。

在这个例子中，我们考虑一个更为复杂的阈值处理问题。图 10.43(a)显示了一幅酵母细胞的 8 比特图像，其中，我们希望用全局阈值处理得到相对于亮点的区域。首先，图 10.43(b)显示了该图像的直方图，图 10.43(c)是把 Otsu 方法直接用于所示直方图的图像上得到的结果。我们看到，Otsu 方法未能实现检测亮点这一原始目标，并且，尽管该方法能隔离出某些细胞区域本身，但右侧的几个分割后的区域却未分开。Otsu 方法计算出的阈值是 42，可分性测度是 0.636。

图 10.43(d)显示了图像 $g_T(x, y)$，该图像是按如下方式得到的：先计算拉普拉斯图像的绝对值，然后用区间[0, 255]内的灰度级 115 作为 T 值进行阈值处理。这个 T 值近似对应于拉普拉斯图像绝对值中的 99.5%，所以在这个水平上阈值处理应该得到一个稀疏的像素集合，如图 10.43(d)所示。注意，在该图像中，很多点聚集在亮点的边缘附近，正如前面的讨论所预料的那样。图 10.43(e)是图(a)和图(d)之积的非零像素的直方图。最后，图 10.43(f)显示了以图 10.43(e)中的直方图为基础，采用 Otsu 方法全局分割原图像后的结果。该结果与图像中亮点的位置一致。用 Otsu 方法计算出的阈值是 115，可分性测度是 0.762，这两个值都要高于使用原始直方图得到的值。

772～774

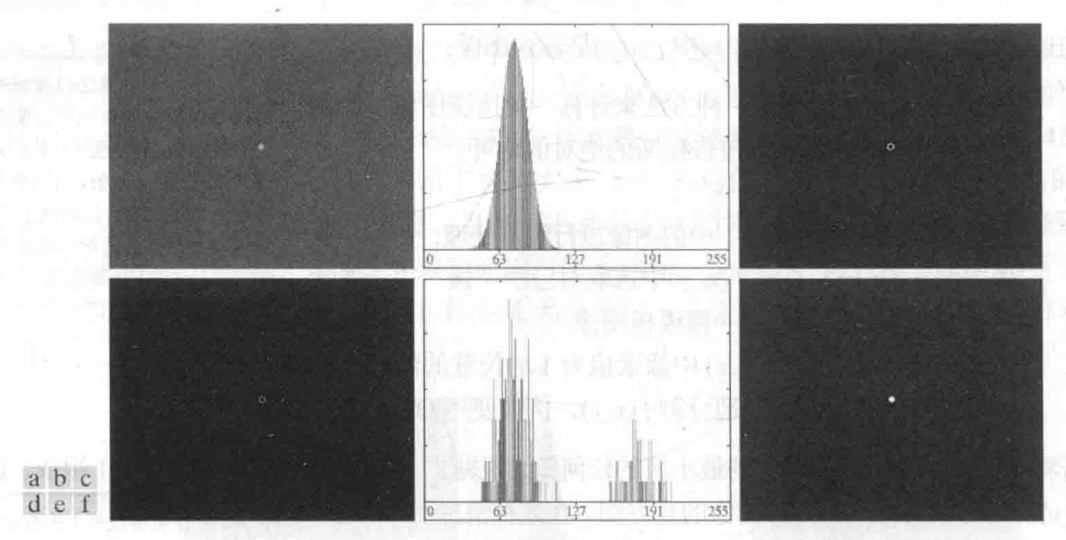


图 10.42 (a)来自图 10.41(a)的带噪声图像；(b)该图像的直方图；(c)在 99.7%处经阈值处理后的梯度幅度图像；(d)图(a)和图(c)之积形成的图像；(e)图(d)所示图像中非零像素的直方图；(f) 以直方图(e)为基础，用 Otsu 阈值处理分割图像的结果。阈值为 134，在该直方图中处在两个波峰间的近似中间的位置

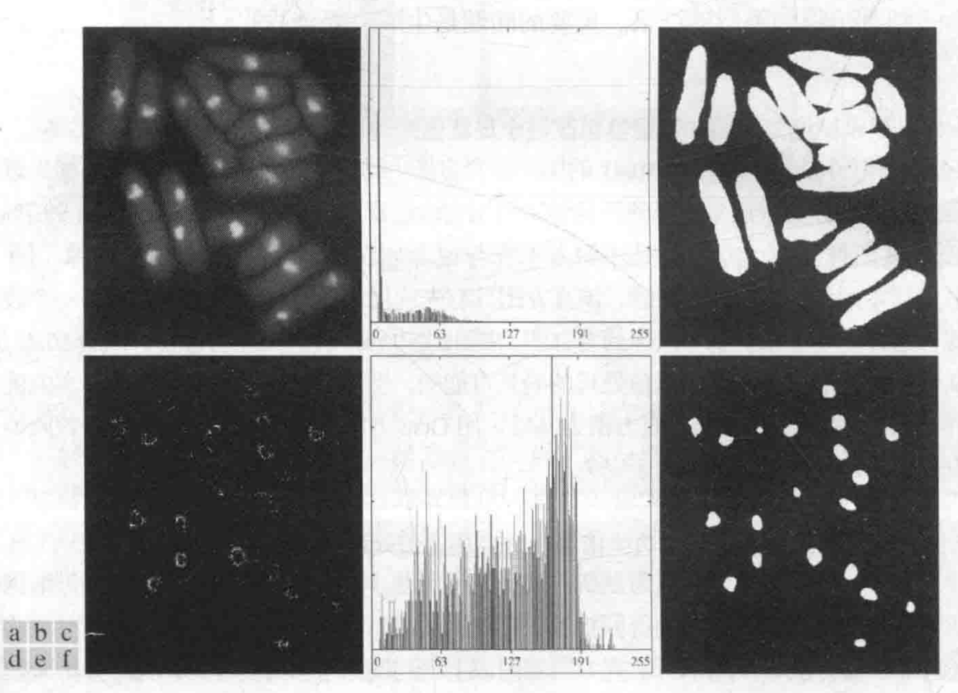


图 10.43 (a)酵母细胞图像；(b)图(a)的直方图；(c)使用图(b)中的直方图，用 Otsu 方法对图(a)进行分割；(d)经阈值处理后的绝对拉普拉斯图像；(e)图(a)和图(d)乘积中的非零像素的直方图；(f)以(e)中的直方图为基础，使用 Otsu 的方法对原图像进行阈值处理后的结果(原图像由南加利福尼亚大学的 Susan L. Forsburg 教授提供)

在不同百分比处设置阈值，我们甚至还可以改善细胞区域的分割。例如，图 10.44 显示了使用前段中的相同步骤得到的结果，但阈值设置在 55 处，它近似拉普拉斯图像最大绝对值的 5%。该值位于图像中的所有值的 53.9%。很明显，这个结果好于图 10.43(c)所示对原始图像的直方图用 Otsu 的方法得到的结果。

774 ~ 775

10.3.6　多阈值处理

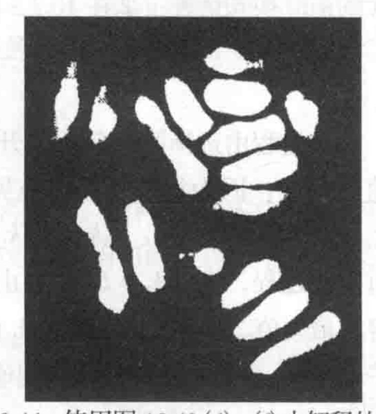


图 10.44　使用图 10.43(d)~(f) 中解释的相同过程，对图 10.43(a) 中的图像进行分割后的结果，但对绝对拉普拉斯图像进行阈值处理时，使用了一个较低的阈值

迄今为止，我们关注的是用单个全局阈值对图像进行分割。我们可将 10.3.3 节中介绍的阈值处理方法扩展到任意数量的阈值，因为以这种方法为基础的可分性测度也可以扩展到任意数量的分类(Fukunaga[1972])。在 K 个类 $C_1, C_2, \cdots, C_K$ 的情况下，类间方差可归纳为下面的表达式：

$$\sigma_B^2 = \sum_{k=1}^{K} P_k (m_k - m_G)^2 \tag{10.3-21}$$

式中，

$$P_k = \sum_{i \in C_k} p_i \tag{10.3-22}$$

$$m_k = \frac{1}{P_k} \sum_{i \in C_k} i p_i \tag{10.3-23}$$

并且，m_G 是由式(10.3-9)给出的全局均值。K 类由 $K-1$ 个阈值来分离，这些值 $k_1^*, k_2^* \cdots, k_{K-1}^*$ 是式(10.3-21)的最大值：

$$\sigma_B^2(k_1^*, k_2^*, \cdots, k_{K-1}^*) = \max_{0<k_1<k_2<\cdots k_{n-1}<L-1} \sigma_B^2(k_1, k_2, \cdots, k_{K-1}) \tag{10.3-24}$$

虽然这个结果通常很完美，但当分类数量增加时它会开始失去意义，因为我们仅处理一个变量(灰度)。事实上，通常类间方差是依据以向量描述的多个变量来计算的(Fukunaga[1972])。实际中，当我们有理由相信使用两个阈值可以有效地解决问题时，那么使用多个全局阈值处理就被视为一种可行的方法。要求两个以上阈值的应用，通常使用更多的灰度值来解决。而该方法使用的是附加的描述子(譬如彩色)，并且这种应用是以如 10.3.8 节描述的模式识别问题来筹划的。

对于由三个灰度间隔组成的三个类(这三个类由两个阈值分隔)，类间方差由下式给出：

> 带有两个阈值的阈值处理有时也称为滞后阈值处理。

$$\sigma_B^2 = P_1(m_1 - m_G)^2 + P_2(m_2 - m_G)^2 + P_3(m_3 - m_G)^2 \tag{10.3-25}$$

式中，

$$P_1 = \sum_{i=0}^{k_1} p_i, \quad P_2 = \sum_{i=k_1+1}^{k_2} p_i, \quad P_3 = \sum_{i=k_2+1}^{L-1} p_i \tag{10.3-26}$$

和

$$m_1 = \frac{1}{P_1}\sum_{i=0}^{k_1} i p_i, \quad m_2 = \frac{1}{P_2}\sum_{i=k_1+1}^{k_2} i p_i, \quad m_3 = \frac{1}{P_3}\sum_{i=k_2+1}^{L-1} i p_i \tag{10.3-27}$$

如在式(10.3-10)和式(10.3-11)中那样，如下关系成立：

$$P_1 m_1 + P_2 m_2 + P_3 m_3 = m_G \tag{10.3-28}$$

和

$$P_1 + P_2 + P_3 = 1 \tag{10.3-29}$$

774 ~ 777

因此，我们看到，P 项、m 项及 σ_B^2 都是 k_1 和 k_2 的函数。两个最佳阈值 k_1^* 和 k_2^* 是使得 $\sigma_B^2(k_1, k_2)$ 最

大的值。换句话说，如在 10.3.3 节讨论的单阈值情况那样，我们用下式寻找最佳阈值：

$$\sigma_B^2(k_1^*, k_2^*) = \max_{0<k_1<k_2<L-1} \sigma_B^2(k_1, k_2) \tag{10.3-30}$$

该过程由选择第一个 k_1 值开始（该值是 1，因为在 0 灰度处寻找阈值没有意义；还要记住，增量值为整数，因为我们处理的是灰度）。接着，k_2 的所有值在大于 k_1 和小于 $L-1$ 的范围内增加(即 $k_2 = k_1+1,\cdots,L-2$)。然后，将 k_1 增大到其下一个值，k_2 的所有值再次在大于 k_1 的所有值范围内增加。重复该过程，直到 $k_1 = L-3$ 为止。该处理的结果是一个二维阵列 $\sigma_B^2(k_1,k_2)$，最后一步是在该阵列中寻找最大值。对应于最大值的 k_1 值和 k_2 值就是最佳阈值 k_1^* 和 k_2^*。如果存在几个最大值，则平均对应于 k_1 和 k_2 的值来得到最终的阈值。然后，阈值处理后的图像由下式给出：

$$g(x, y) = \begin{cases} a, & f(x, y) \leqslant k_1^* \\ b, & k_1^* < f(x, y) \leqslant k_2^* \\ c, & f(x, y) > k_2^* \end{cases} \tag{10.3-31}$$

式中，a, b 和 c 是任意三个有效的灰度值。

最后，我们注意到，在 10.3.3 节中为单个阈值定义的可分性测度可直接扩展到多个阈值：

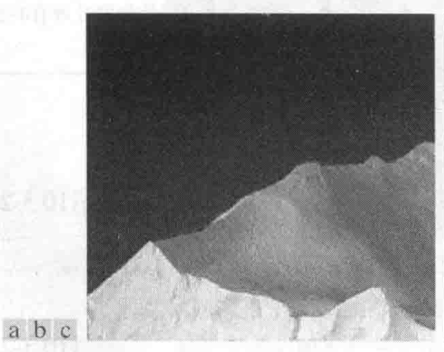
$$\eta(k_1^*, k_2^*) = \frac{\sigma_B^2(k_1^*, k_2^*)}{\sigma_G^2} \tag{10.3-32}$$

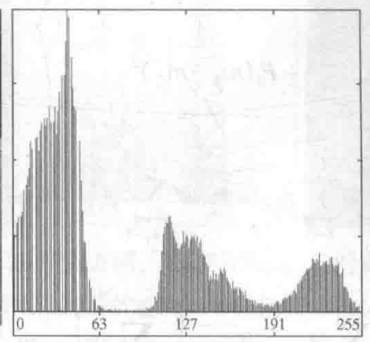
式中，σ_G^2 是来自式(10.3-13)的总图像方差。

例 10.19 多个全局阈值处理。

图 10.45(a)显示了一幅冰山的图像。这个例子的目的是把图像分割成三个区域：暗背景、冰山的明亮区域和阴影区域。由图 10.45(b)中的图像直方图可以看出，解决这一问题需要有两个阈值。按上面讨论的过程得到阈值 $k_1^* = 80$ 和 $k_2^* = 177$，由图 10.45(b)注意到，它们靠近两个直方图波谷的中心。图 10.45(c)是在式(10.3-31)中使用这两个阈值得到的分割。可分性测度是 0.954。这个例子能做到这样好的主要原因是，能寻找到三个具有明显可分模式的直方图，直方图具有适度宽度和深度的波谷。

a b c





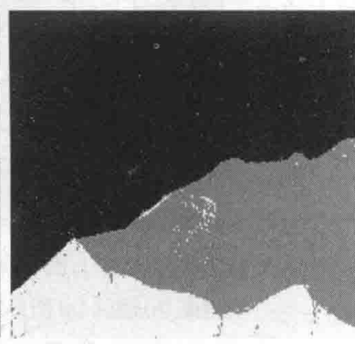


图 10.45 (a)冰山图像；(b)直方图；(c)用双 Otsu 阈值将图像分割为三个区域（原图像由 NOAA 提供）

10.3.7 可变阈值处理

777 ~ 778

正如 10.3.1 节中讨论的那样，噪声和非均匀光照这样的因素对阈值处理算法的性能起着重要作用。我们在 10.3.4 节和 10.3.5 节说过，图像平滑和边缘信息的使用有益于阈值处理。然而，常

常会出现这种情况，即用迄今为止讨论的任何方法解决问题时，要么这种类型的预处理不切实际，要么对这一情形的改进简单而无效。在这种情况下，更为高级且复杂的阈值处理涉及可变阈值问题。本节讨论选择可变阈值的各种技术。

图像分块

可变阈值处理最简单的方法之一是，把一幅图像分成不重叠的矩形。这种方法用于补偿光照和/或反射的不均匀性。选择的矩形要足够小，以便每个矩形的光照都近似是均匀的。我们用一个例子来说明这一方法。

例 10.20　通过图像分块的可变阈值处理。

图 10.46(a) 显示了来自图 10.37(c) 的图像，图 10.46(b) 显示了其直方图。讨论图 10.37(c) 时，我们可得出这样的结论：使用一个全局阈值不能分割该图像，图 10.46(c) 和 (d) 已确认了这一事实，图 10.46(c) 和 (d) 分别显示了使用 10.3.2 节中讨论的迭代方案和 Otsu 方法分割图像后的结果。两种方法产生了可比较的结果，结果中大量的分割错误显而易见。

图 10.46(e) 显示了已细分为 6 个矩形区域的原图像，图 10.46(f) 是对每幅子图像应用 Otsu 全局方法得到的结果。尽管分割结果中仍然可见一些错误，但在一幅分割起来十分困难的图像上，图像细分产生了合理的结果。分析每幅子图像的直方图，很容易就可解释这种改进的原因。如图 10.47 所示，每幅子图像都由一个双模式直方图来表征，两个模式之间有一个较深的波谷，这是我们所知的将导致有效全局阈值处理的一个事实。

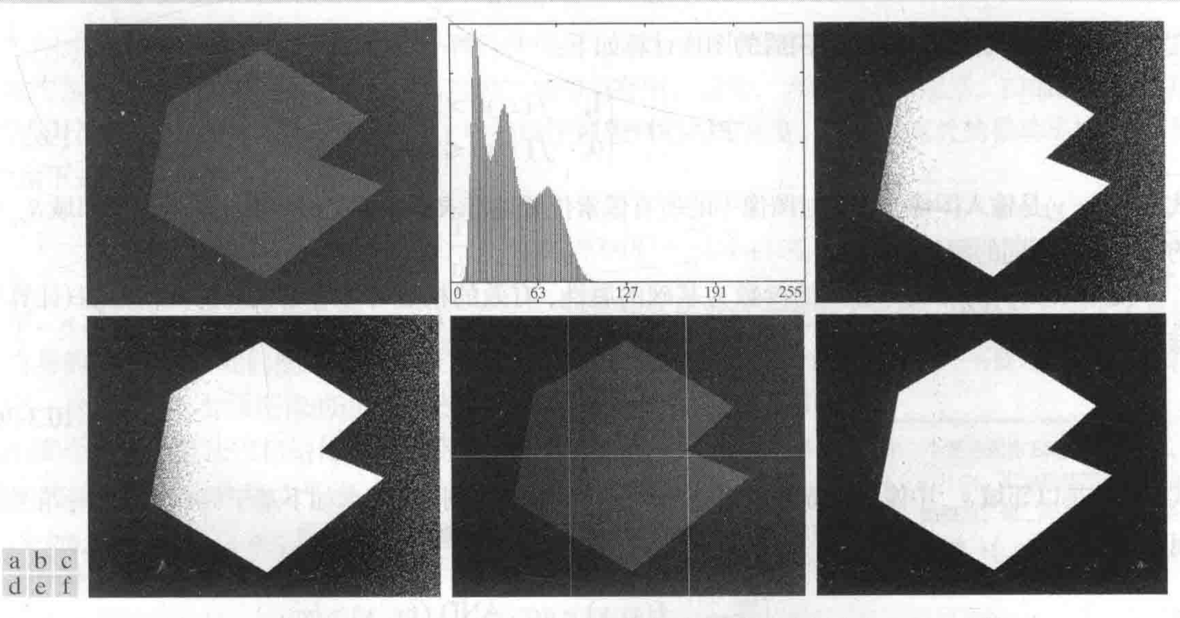


图 10.46　(a) 带噪声阴影的图像；(b) 该图像的直方图；(c) 用来自 10.3.2 节的全局迭代算法对图 (a) 的分割；(d) 使用 Otsu 方法得到的结果；(e) 分为 6 幅子图像后的图像；(f) 分别对每幅子图像应用 Otsu 方法的结果

当感兴趣物体和背景占据合理的可比大小的区域时，通常图像细分工作得较好，如图 10.46 所示。当不是这种情况时，该方法通常会失败，因为细分可能只包含物体或背景像素。虽然使用其他技术来确定何时一个细分包含两种像素可以解决这一情况，但解决不同情况所要求的逻辑会变得很复杂。在这种情况下，使用这一节其余部分讨论的方法可能更好。

778 ~ 780

基于局部图像特性的可变阈值处理

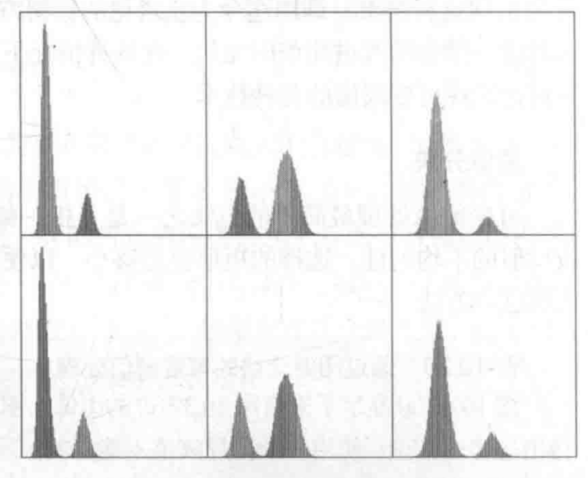


图 10.47 图 10.46(e)中 6 幅子图像的直方图

与前一节中讨论的图像细分方法相比，更为一般的方法是在一幅图像中的每点(x, y)计算阈值，该阈值以一个或多个在(x, y)邻域计算的特性为基础。虽然这可能是有点费力的处理，但现代算法和硬件允许进行快速的邻域处理，特别是对于像逻辑操作和算术操作这样的普通函数。

我们用一幅图像中每个点的一个邻域内的像素的标准差和均值，来说明局部阈值处理的基本方法。这两个量对于确定局部阈值非常有用，因为它们是局部对比度和平均灰度的描述子。令σ_{xy}和m_{xy}表示一幅图像中以坐标(x, y)为中心的邻域S_{xy}所包含的像素集合的标准差和均值(关于局部均值和标准差的计算，见 3.3.4 节)。下面是可变局部阈值的通用形式：

$$T_{xy} = a\sigma_{xy} + bm_{xy} \tag{10.3-33}$$

式中，a 和 b 是非负常数，且

$$T_{xy} = a\sigma_{xy} + bm_G \tag{10.3-34}$$

其中，m_G 是全局图像均值。分割后的图像计算如下：

$$g(x, y) = \begin{cases} 1, & f(x,y) > T_{xy} \\ 0, & f(x,y) \leqslant T_{xy} \end{cases} \tag{10.3-35}$$

式中，$f(x, y)$是输入图像。该式对图像中的所有像素位置进行求值，并在每个点(x, y)处使用邻域S_{xy}中的像素计算不同的阈值。

使用以(x, y)的邻域计算出的参数为基础的属性，有效的权值可显著增强局部阈值处理(计算量会适度增加)：

$$g(x, y) = \begin{cases} 1, & Q\,(\text{局部参数})\text{为真} \\ 0, & Q\,(\text{局部参数})\text{为假} \end{cases} \tag{10.3-36}$$

式中，Q 是以邻域 s_{xy} 中像素计算的参数为基础的一个属性。例如，考虑如下基于局部均值和标准差的属性 $Q(\sigma_{xy}, m_{xy})$：

$$Q(\sigma_{xy}, m_{xy}) = \begin{cases} \text{真}, & f(x, y) > a\sigma_{xy}\,\text{AND}\, f(x, y) > bm_{xy} \\ \text{假}, & \text{其他} \end{cases} \tag{10.3-37}$$

注意，式(10.3-35)是式(10.3-36)的一种特殊情况，它是在 $f(x,y) > T_{xy}$ 时令 Q 为真，而在其他情况下令 Q 为假得到的。在这种情况下，该属性仅简单地基于一个点处的灰度。

例 10.21 基于局部图像特性的可变阈值处理。

图 10.48(a)显示了来自例 10.18 的酵母图像。该图像有三个主要的灰度级，因此双阈值处理是一种较好分割方法的假设是合理的。图 10.48(b)是使用 10.3.6 节中说明的双阈值处理方法得到的结果。如图所示，从背景中分离出明亮区域是可能的，但图像右侧的中等灰度区域未完全分割［回忆我们在

780 ~ 781

例 10.18 中遇到过的与图 10.43(c)类似的问题]。为说明局部阈值处理的用法，我们用大小为 3×3 的邻域对输入图像中的所有 (x, y) 计算局部标准差 σ_{xy}。图 10.48(c)显示了结果。注意，微弱的外部线条是如何正确地表征细胞的边界的。接着，我们使用全局均值替代 m_{xy} 来形成式(10.3-37)所示形式的一个属性。通常，当背景近似恒定，且所有物体的灰度高于背景灰度或低于背景灰度时，选择全局均值会给出较好的结果。在完成的属性规范中，使用了值 $a = 30$ 和 $b = 1.5$（在类似这样的应用中，这些值通常是通过试验确定的）。然后，用式(10.3-36)来分割图像。如图 10.48(d)所示，结果与输入图像中的两类灰度的大部分区域十分接近。尤其要注意所有的外部区域完全被正确分割，且大多数较亮的内部区域也已被正确地分开。

a b c d

图 10.48　(a)来自图 10.43 的图像；(b)用 10.3.6 节中讨论的双阈值处理方法分割的图像；(c)局部标准差图像；(d)用局部阈值处理得到的结果

使用移动平均

刚才讨论的局部阈值处理方法的一种特殊情形，是以一幅图像的扫描行计算移动平均为基础的。在速度是一种基本要求的文档处理中，这种实现相当有用。通常，为减少光照偏差，扫描是以 Z 字形模式逐行执行的。令 z_{k+1} 表示步骤 $k+1$ 中扫描序列遇到的点的灰度。这个新点处的移动平均（平均灰度）由下式给出：

$$m(k+1) = \frac{1}{n}\sum_{i=k+2-n}^{k+1} z_i = m(k) + \frac{1}{n}(z_{k+1} - z_{k-n}) \tag{10.3-38}$$

式中，n 是用于计算平均的点数，$m(1) = z_1/n$。这个初始值严格来说并不正确，因为单点的平均是该点自身的值。然而，我们使用 $m(1) = z_1/n$，因此第一次开始计算式(10.3-38)时，并不要求特殊的计算。从另一个角度看，如果图像的边界填充了 $n-1$ 个 0，它就是我们将得到的那个值。该算法仅初始化一次，而不在每行都初始化。因为对图像中的每个点都计算移动平均，因此用式(10.3-35)实现分割，其中 $T_{xy} = bm_{xy}$，b 是常数，m_{xy} 是在输入图像中的点(x, y)处使用式(10.3-38)得到的移动平均。

> 第一个表达式在 $k \geq n-1$ 时有效。当 $k < n-1$ 时，平均是使用可用点形成的。类似地，第二个表达式在 $k \geq n+1$ 时有效。

例 10.22　用移动平均的文档阈值处理。

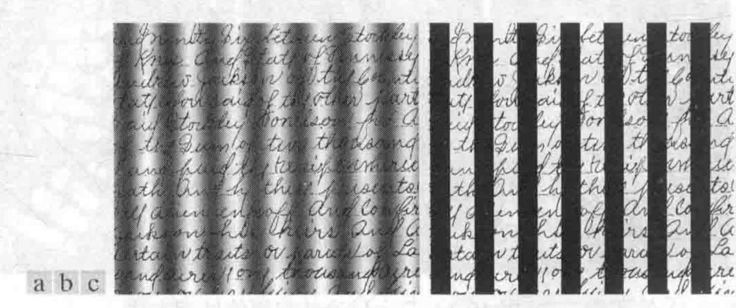
图 10.49(a)显示了一幅以斑点灰度模式遮蔽的手写文本图像。这种灰度遮蔽形式是使用摄影闪光灯得到的典型图像。图 10.49(b)是使用 Otsu 全局阈值处理方法分割的结果。全局阈值处理不能克服灰度变化并不意外。图 10.49(c)显示了采用移动平均的局部阈值处理成功分割的结果。经验之一是令 n 等于平均笔画宽度的 5 倍。在这种情况下，平均宽度是 4 个像素，因此我们在式(10.3-38)中令 $n = 20$ 并使用 $b = 0.5$。

作为这种分割方法有效性的另一个说明，我们使用了与前一段中相同的参数来分割图 10.50(a)中的图像，该图像被正弦型灰度变化污染，文本扫描仪的电源未适当接地时，会出现这种典型的变化。如图 10.50(b)和(c)所示，分割结果类似于图 10.49 中的那些结果。

781～783

a b c

图 10.49　(a)被斑点遮蔽污染的文本图像；(b)使用 Otsu 方法的全局阈值处理的结果；(c)使用移动平均的局部阈值处理的结果



a b c

图 10.50　(a)被正弦亮度遮蔽污染的文本图像；(b)使用 Otsu 方法的全局阈值处理的结果；(c)使用移动平均的局部阈值处理的结果

有趣的是，使用相同的 n 值和 b 值，两种情形下均得到了成功的分割结果，这说明了这种方法的相对稳定性。通常，当感兴趣的物体与图像尺寸相比较小(或较细)时，基于移动平均的阈值处理会工作得很好。打印图像和手写文本图像满足这一条件。

10.3.8　多变量阈值处理

迄今为止，我们已涉及了基于单个变量(灰度)的阈值处理。在某些情况下，传感器可产生多个可利用的变量来表征图像中的每个像素，这样，就允许进行多变量阈值处理。一个显著的例子就是彩色成像，其中红(R)、绿(G)和蓝(B)分量用于形成一幅合成彩色图像(见第 6 章)。在这种情况下，每个"像素"由三个值来表征，并且可以表示为一个三维向量 $\mathbf{z}=(z_1,z_2,z_3)^{\mathrm{T}}$，其分量是一个点的 RGB 彩色。这些三维点通常称为体素，以便与图像元素相对应来表示体积元素。

如 6.7 节中讨论的某些细节那样，多变量阈值处理可视为一种距离计算。假设我们想要从一幅彩色图像中提取具有指定颜色的所有区域，比如说，微红色区域。令 $\mathbf{a}$ 表示我们感兴趣的平均微红色。基于这一参数来分割彩色图像的一种方法是，计算任意彩色点 $\mathbf{z}$ 和平均彩色 $\mathbf{a}$ 间的距离测度 $D(\mathbf{z},\mathbf{a})$。然后，我们按如下方式分割输入图像：

$$g=\begin{cases}1, & D(\mathbf{z},\mathbf{a})<T\\ 0, & \text{其他}\end{cases}\tag{10.3-39}$$

其中，T 是一个阈值。当然，要在输入图像中的所有坐标上执行距离计算，以便产生相应的分割值 g。注意，该式中的不等式与我们在式(10.3-1)中对单个阈值处理所用的不等式相反，原因是等式 $D(\mathbf{z},\mathbf{a})=T$ 定义了一个容积(见图 6.43)，并且它更直观地把分割后的像素值想象为包含在一个容积中，而使背景像素值位于该容积的表面或外部。令 $D(\mathbf{z},\mathbf{a})=-f(x,y)$，式(10.3-39)便可简化为式(10.3-1)。

783～784

很明显，条件$f(x, y)>T$基本上表示f的值和实线的原点间的欧氏距离超过了T值。这样，阈值处理就是基于距离测度的计算，并且式(10.3-39)的形式依赖于所用的测度。通常，如果z是一个n维向量，我们从2.6.6节可知n维欧几里得距离定义为

$$D(z, \boldsymbol{a}) = \|z-\boldsymbol{a}\| = \left[(z-\boldsymbol{a})^{\mathrm{T}}(z-\boldsymbol{a})\right]^{\frac{1}{2}} \tag{10.3-40}$$

等式$D(z, \boldsymbol{a}) = T$描述了欧几里得空间中(图6.4.3中显示了一个三维例子)的一个球(称为超球)。一种更有用的距离测度是所谓的马氏(Mahalanobis)距离，它定义为

$$D(z, \boldsymbol{a}) = \left[(z-\boldsymbol{a})^{\mathrm{T}}\boldsymbol{C}^{-1}(z-\boldsymbol{a})\right]^{\frac{1}{2}} \tag{10.3-41}$$

其中，如12.2.2节讨论的那样，$\boldsymbol{C}$是z的协方差矩阵。$D(z, \boldsymbol{a}) = T$描述了一个n维超球(图6.43显示了一个三维例子)。当$\boldsymbol{C} = \boldsymbol{I}$时，即它是么阵时，该表达式就简化为式(10.3-40)。

我们在6.7节中已给出了一个关于使用这些表达式的详细例子。我们在12.2节中还要讨论使用基于决策函数的模式识别技术来分割图像的问题，这种问题可视为多类、多变量阈值处理问题。

10.4 基于区域的分割

如10.1节中讨论的那样，分割的目的是将一幅图像划分为多个区域。在10.2节中，我们其于灰度级的不连续性尝试寻找区域间的边界来解决这一问题；而在10.3节中，分割是通过以像素特性分布为基础的阈值处理来完成的，如灰度值或彩色。本节讨论以直接寻找区域为基础的分割技术。

> 在继续阅读下文前，建议读者回顾一下10.1节中介绍的术语。

10.4.1 区域生长

如其名字所示的那样，区域生长是根据预先定义的生长准则，将像素或子区域组合为更大区域的过程。基本方法是从一组“种子”点开始，将与种子预先定义的性质相似的那些邻域像素添加到每个种子上，来形成这些生长区域(如特定范围的灰度或颜色)。

如稍后在例10.23中显示的那样，通常我们可基于问题的性质来选择一组或多组开始点。当先验知识不可用时，这一过程是在每个像素处计算相同特性的集合，最终在生长过程期间，根据这些特性集合把像素分配给各个区域。如果这些计算的结果显示了一族值，则那些特性靠近这些族的中心的像素可以作为种子使用。

相似性准则的选择不仅取决于所面对的问题，还取决于现有的图像数据类型。例如，土地利用卫星成像分析非常倚重彩色的使用。如果彩色图像中没有可用的固有信息，那么解决这个问题将非常困难，甚至无法解决。如果图像是单色的，就须使用一组基于灰度级和空间特性的描绘子(如矩或纹理)对区域进行分析。在第11章中，我们将讨论那些对区域特征有用的描绘子。

如果在区域生长过程中未使用连通属性，那么单独的描绘子会产生错误的结果。例如，设想仅有三个不同灰度值的像素的随机排列。将具有相同灰度级的像素组合形成一个“区域”，而不考虑连通性，会产生对当前讨论内容毫无意义的分割结果。

区域生长的另一个问题是终止规则的表示法。当不再有像素满足加入某个区域的准则时，区域生长就会停止。像灰度值、纹理和彩色准则本质上都是局部的，都没有考虑区域生长的“历史”。增强区域生长算法能力的其他准则，利用了候选像素和已加入生长区域的像素间的大小、相似性等概念(如

784 ~ 786

候选像素灰度和生长区域的平均灰度的比较)，以及正在生长的区域的形状。这类描绘子的使用是以期望结果的模型至少部分可用这一假设为基础的。

令$f(x, y)$表示一个输入图像阵列；$S(x, y)$表示一个种子阵列，阵列中种子点位置处为1，其他位置处为0；Q表示在每个位置(x, y)处所用的属性。假设阵列f和S的尺寸相同。基于8连接的一个基本区域生长算法可说明如下。

关于连通分量，请参阅2.5.2节和9.5.3节；关于腐蚀，请参阅9.2.1节。

1. 在$S(x, y)$中寻找所有连通分量，并把每个连通分量腐蚀为一个像素；把找到的所有这种像素标记为1，把S中的所有其他像素标记为0。
2. 在坐标对(x, y)处形成图像f_Q：若输入图像在该坐标处满足给定的属性Q，则令$f_Q(x, y) = 1$，否则令$f_Q(x, y) = 0$。
3. 令g是这样形成的图像：即把f_Q中为8连通种子点的所有1值点，添加到S中的每个种子点。
4. 用不同的区域标记(如1, 2, 3,…)标出g中的每个连通分量。这就是由区域生长得到的分割图像。

我们通过一个例子来说明这种算法的机理。

例10.23　使用区域生长的分割。

图10.51(a)显示了一条焊缝(水平深色区域)的8比特X射线图像，该图像中含有几条裂缝和孔隙(水平横穿图像中心的明亮区域)。我们通过分割有缺陷的焊缝区域来说明区域生长的应用。这些区域可用于焊缝检测，包括有关历史学习的数据库或控制自动焊接系统。

工作的第一步是确定种子点。从问题的物理性质看，我们知道裂缝和孔隙与实焊缝相比，对X射线的衰减要小，因此，我们预料包含这种类型缺陷的区域要比X射线图像中的其他部分明显亮一些。我们可用在高百分比处设置阈值对原始图像进行阈值处理来提取种子点。图10.51(b)显示了图像的直方图，图10.51(c)显示了用等于图像中灰度值的99.5%的阈值对图像进行阈值处理后的结果，在这种情况下，阈值为254(见10.3.5节关于百分比的说明)。图10.51(d)显示了将图10.51(c)中的每个连通分量用形态学腐蚀为一个单点后的结果。

接下来，我们必须确定一个属性。在这个例子中，我们的兴趣是把满足如下要求的所有像素添加到每个种子中：(a)对于种子是8连接的，并且(b)与种子相似。若使用灰度差作为一种相似性测度，则应用于每个位置(x, y)处的属性是

$$Q = \begin{cases} \text{TRUE}, & \text{种子和点}(x, y)\text{处的像素间灰度差的绝对值} \leqslant T \\ \text{FALSE}, & \text{其他} \end{cases}$$

其中，T是一个指定的阈值。虽然这个属性是基于灰度差的，并且使用单个阈值，但我们可指定更复杂的方案，即对每个像素应用一个不同的阈值，并使用不同于差值的其他属性。在这种情况下，前述的属性足以解决该问题，如这个例子的其余部分说明的那样。

从前面的章节我们知道，最小的种子值是255，因此，使用254的阈值对图像进行阈值处理。图10.51(e)显示了图10.51(a)和(c)所示图像的差的绝对值。图10.51(e)中的图像包含了用于计算每个位置(x, y)处的属性所需要的差值。图10.51(f)显示了相应的直方图。我们需要一个阈值，以便用于建立相似性属性。该直方图有三个主要模式，因此我们首先把10.3.6节中讨论的双阈值处理技术用于差值图像。在这种情况下，导致的两个阈值是$T_1 = 68$和$T_2 = 126$，我们看到，它们相当靠近直方图的波谷［暂离开这一主题，我们用这两个阈值分割了图像。图10.51(g)显示了使用双阈值不能解决缺陷的分割问题，尽管阈值处在主要波谷中］。

786
~
788

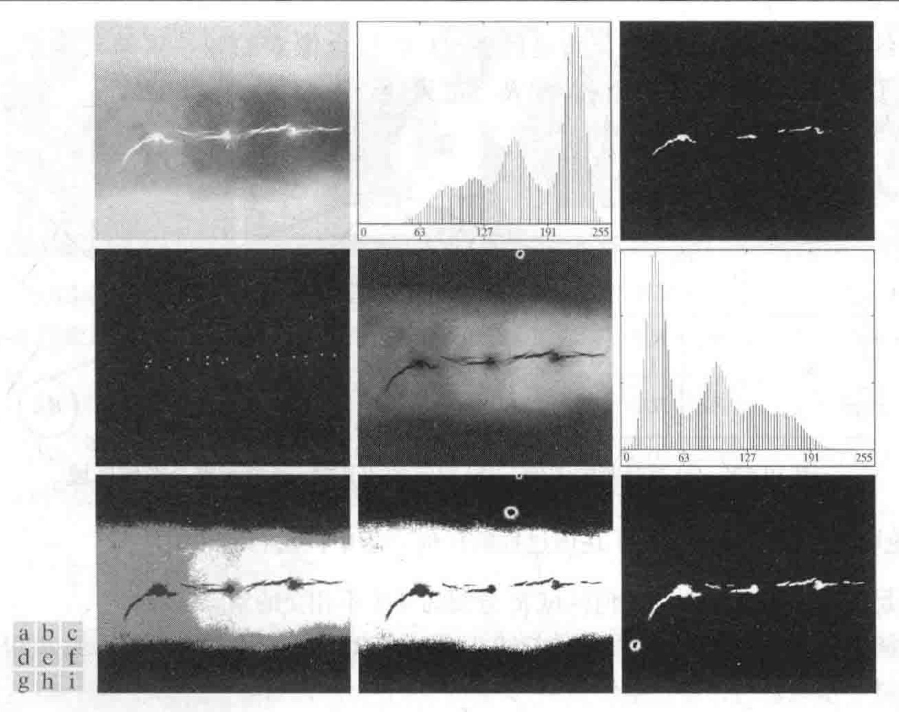


图 10.51 (a)有缺陷焊缝的X射线图像；(b)直方图；(c)初始种子图像；(d)最终种子图像(为清楚起见，点已被放大)；(e)图(a)和图(c)之差的绝对值；(f)图(e)的直方图；(g)使用双阈值进行阈值处理后的差值图像；(h)使用最小双阈值进行阈值处理后的差值图像。(i)用区域生长得到的分割(原图像由X-TEK Systems有限公司提供)

图10.51(h)显示了仅使用T_1对差值图像进行阈值处理的结果。黑点是其属性为“真”(TRUE)的像素，其他点是属性为“假”(FALSE)的像素。这里，重要结果是焊缝的良好区域中断定为失败的点，因此，这些点将不包含在最终结果中。外部区域的点将由区域生长算法作为候选点来考虑。然而，步骤3将拒绝外部点，因为它们与种子不是8连接的。事实上，如图10.51(i)所示，这一步导致了正确的分割，这表明连接性的使用是这种情况下的一个基本需求。最后，注意，在步骤4中，我们对由该算法找到的所有区域使用了相同的值。在这种情况下，它确实做得很好。

10.4.2 区域分裂与聚合

上一节中讨论的过程是从一组种子点来生长区域。另一种可供选择的方法是，首先将一幅图像细分为一组任意的不相交区域，然后聚合和/或分裂这些区域，并试图满足10.1节中阐述的分割条件。下面讨论分裂和聚合的基础知识。

令R表示整幅图像区域，并选择一个属性Q。对R进行分割的一种方法是，依次将它细分为越来越小的四象限区域，以便对于任何区域R_i有$Q(R_i)=\text{TRUE}$。我们从整个区域开始。如果$Q(R)=\text{FALSE}$，那么将该图像分割为4个象限区域。如果对于每个象限区域Q为FALSE，则将该象限区域再次细分为四个子象限区域，以此类推。这种特殊的分裂技术有一种方便的表示方法，即用所谓的四叉树形式表示，也就是每个节点都正好有4个后代，如图10.52所示(对应一个四叉树的节点的图像有时称为*四分区域*或*四分图像*)。注意，树根对应于整幅图像，而每个节点对应于该节点的4个细分子节点。在这种情况下，仅R_4被进一步细分。

如果只使用分裂，那么最后的分区通常包含具有相同性质的邻接区域。这种缺陷可以通过允许聚合和分裂得到补救。要满足10.1节中

> 关于区域相邻的讨论，请参阅2.5.2节。

787
~
789

提出的分割约束条件，仅要求聚合满足属性 Q 的组合像素的邻接区域。也就是说，只有在 $Q(R_j \cup R_k) = \text{TRUE}$ 时，两个邻接区域 R_j 和 R_k 才能聚合。

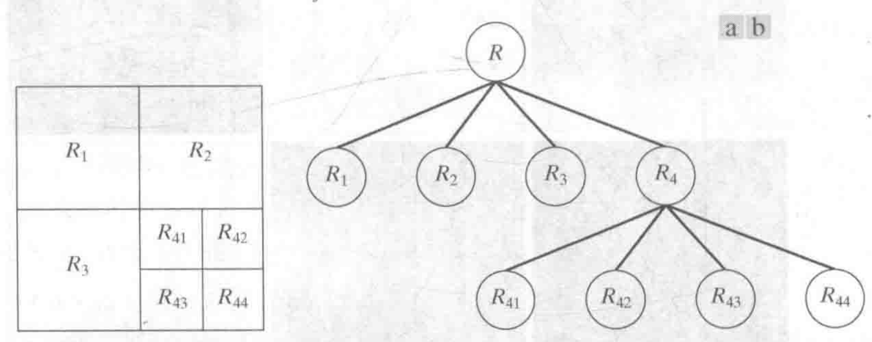


图 10.52 (a)被分割的图像；(b)对应的四叉树。R 表示整个图像区域

前述讨论可以小结为如下过程，在该过程的任何一步中，我们

1. 把满足 $Q(R_i) = \text{FALSE}$ 的任何区域 R_i 分裂为 4 个不相交的象限区域。
2. 不可能进一步分裂时，对满足条件 $Q(R_j \cup R_k) = \text{TRUE}$ 的任意两个邻接区域 R_j 和 R_k 进行聚合。
3. 无法进一步聚合时，停止操作。

习惯上要规定一个不能再进一步执行分裂的最小四象限区的尺寸。

前述的基本原理可能有多种变化。例如，在步骤 2 中，如果两个邻接的区域 R_i 和 R_j 各自都满足属性，则我们允许聚合这两个区域会得到有意义的简化。这将导致简单得多(且快得多)的算法，因为属性的测试被限制在各个四分区域内。正如下例所示，这种简化仍能产生较好的分割结果。

例 10.24 用区域分裂和聚合进行分割。

图 10.53(a)显示了天鹅星座环的一幅大小为 566×566 的 X 射线波段图像。这个例子的目的是把围绕致密中心的不太密集的环状物质的图像分割出来。感兴趣区域有一些有助于分割的明显特征。首先，我们注意到该区域中的数据具有随机性质，这表明其标准差应大于背景和较大的平滑中心区域的标准差(背景的标准差近于 0)。类似地，来自外环的包含数据的一个区域的均值(平均灰度)应大于暗背景的均值，而小于较大且明亮中心区域的均值。因此，我们应该能够用下面的属性分割感兴趣的区域：

$$Q = \begin{cases} \text{TRUE}, & \sigma > a \text{且} 0 < m < b \\ \text{FALSE}, & \text{其他} \end{cases}$$

式中，m 和 σ 是一个四象限区域中的像素的均值和标准差，a 和 b 为常数。

对感兴趣外部区域中的几个区域的分析表明，这些区域中的像素的平均灰度未超过 125，而标准差总是大于 10。图 10.53(b)到(d)显示了对 a 和 b 使用这些值和四象限区域的最小尺寸从 32 变到 8 得到的结果。一个四象限区域中满足属性的像素置为白色，而该区域中的其他像素则置为黑色。根据截获的外部区域形状，最好的结果是使用大小为 16×16 的四象限区域得到的。图 10.53(d)中的黑色方块是大小为 8×8 的四象限区域，该区域中的像素不满足属性。使用较小的四象限区域将导致此类黑色区域数量的增加。使用比此处说明的大区域会导致更多“类似块”的分割。注意，在所有情况下，分割后的区域(白色像素)从背景中完全分隔了平滑的内部区域。这样，分割就有效地把图像分成了三个明显的区域，这些区域对应于图像中的三个主要特征：背景、密集区域和稀疏区域。若使用图 10.53 中的任何一个白色区域作为模板，会使得从原图像中提取这些区域很简单(见习题 10.40)。如例 10.23 中那样，使用基于边缘或阈值的分割不可能得到这些结果。

789 ~ 791

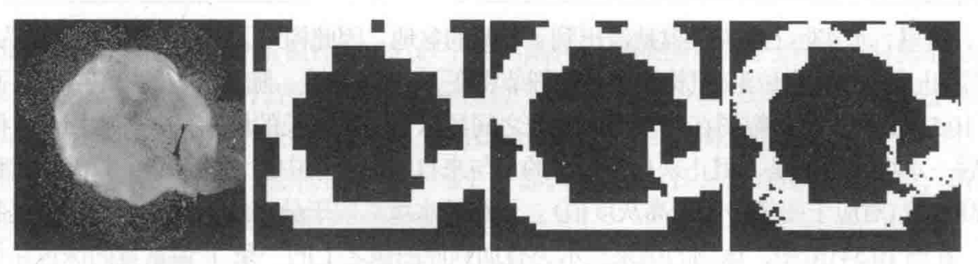



图 10.53　(a) 由 NASA 的哈勃望远镜在 X 射线波段拍摄的天鹅星座环超新星的图像；(b)~(d) 分别将允许的最小四象限区域尺寸限制为 32×32, 16×16 和 8×8 像素时得到的结果(原图像由 NASA 提供)

如前面例子中所用的那样，基于一个区域内像素灰度的均值和标准差的特性，试图量化该区域的纹理（关于纹理的讨论，见 11.3.3 节）。纹理分割的概念是以在属性上使用纹理测度为基础的。换句话说，我们可以使用本节讨论的任何一种方法，用指定基于纹理内容的属性简单地执行纹理分割。

10.5　用形态学分水岭的分割

迄今为止，我们已经讨论了基于三种主要的分割：(a) 边缘检测，(b) 阈值处理和 (c) 区域生长。每种方法都有其优点(例如全局阈值处理具有速度优势) 和缺点(例如，在基于边缘的分割中，需要进行后处理，如边缘连接)。本节讨论基于形态学分水岭概念的方法。在下面的讨论中，形态学分水岭分割将其他三种方法中的许多概念进行了具体化，因此通常会产生更稳定的分割结果，包括连接的分割边界。这种方法为在分割中结合基于知识的约束提供了一个简单的框架。

10.5.1　背景知识

分水岭的概念是以三维方式来形象化一幅图像为基础的：两个空间坐标作为灰度的函数，如图 2.18 (a) 所示。在这种“地形学”解释中，我们考虑三种类型的点：(a) 属于一个区域最小值的点；(b) 把一点视为一个水滴，如果把这些点放在任意位置上，水滴一定会下落到某个最小值点；(c) 处在该点的水会等概率地流向不止一个这样的最小值点。对于一个特定的区域最小值，满足条件 (b) 的点的集合称为该最小值的汇水盆地或分水岭。满足条件 (c) 的点形成地面的峰线，它称为分割线或分水线。

基于这些概念的分割算法的主要目标是找出分水线，其基本思想非常简单，如下面类似的说明那样。假设在每个区域的最小值上打一个洞，并且让水通过洞以均匀的速率上升，从低到高淹没整个地形。当不同汇水盆地中上升的水聚集时，修建一个水坝来阻止这种聚合。水将达到在水线上只能见到各个水坝的顶部的程度。这些大坝的边界对应于分水岭的分割线。因此，它们是由分水岭算法提取出来的 (连接的) 边界。

借助于图 10.54，我们可以进一步解释这些概念。图 10.54 (a) 显示了一幅灰度图像，图 10.54 (b) 是一幅地形俯视图，其中“山峰”的高度与输入图像中的灰度值成正比。为便于解释，这些结构的背面已被遮蔽，以便不会与灰度值混淆，因为我们只对三维表示的普通地形感兴趣。为了阻止上升的水从图像的边缘溢出，我们可以想象整个地形 (图像) 的四周已被比最高山峰更高的水坝包围，最高山峰的值是由输入图像灰度级可能具有的最大值决定的。

假设在每个区域的最小值处打一个洞 [如图 10.54 (b) 中的深色区域所示]，并让水以均匀的速率上升，从低到高淹没整个地形。图 10.54 (c) 说明了被水淹没的第一个阶段，其中显示为浅灰色的“水”仅覆盖了图像中相对较暗的背景区域。在图 10.54 (d) 和 (e) 中，我们看到水现在已分别上升至第一个和第二个汇水盆地。当水继续上升时，最终会从一个汇水盆地溢出到另一个汇水盆地。第一个征兆如图 10.54 (f)

791 ~ 793

所示。这里，水实际上从左侧盆地溢出到了右侧的盆地，因此构筑了一个较短的“水坝”(由几个像素组成)来阻止水在这一水位汇合(构筑水坝的细节将在下一节讨论)。随着水位的不断抬升，更加明显的效果如图 10.54(g)所示。这幅图在两个汇水盆地之间显示了一条更长的坝，另一条水坝在右侧盆地的顶部。构筑后一水坝的目的是，阻止来自盆地中的水与来自对应背景中的水的汇聚。持续该过程，直到达到水的最高水位(对应于图像中的最高灰度值)。最终的水坝对应于分水线，这些分水线就是我们希望的分割结果。在图 10.54(h)中，该例的结果显示为叠加到原图像之上的一条 1 像素宽的深色路径。注意，一条重要的性质就是分水线组成一条连通的路径，这样，在两个区域间就给出了连续的边界。

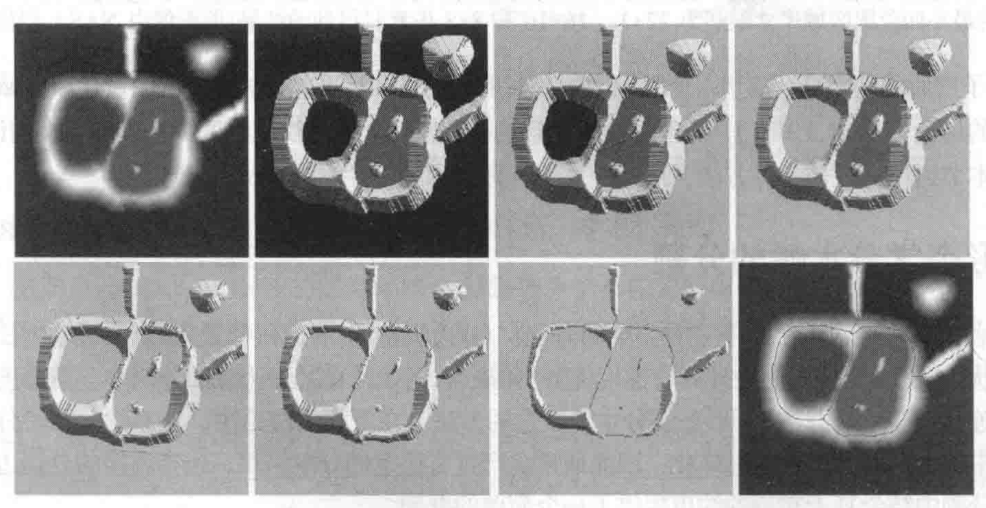


图 10.54 (a)原图像；(b)地形俯视图；(c)~(d)被水淹没的两个阶段；(e)进一步淹没的结果；(f)来自两个汇水盆地的水开始汇聚(两个汇水盆地之间构筑了一个较短的水坝)；(g)较长的水坝；(h)最终的分水(分割)线(原图像由 CMM/Ecole des Mines de Paris 公司的 S. Beucher 博士提供)

分水岭分割的主要应用之一是，从背景中提取近乎一致(类似于水滴)的物体。由变化较小的灰度表征的区域有较小的梯度值。因此，我们经常见到分水岭分割方法用于一幅图像的梯度，而不是图像本身。有了这一表示法，汇水盆地的区域最小值就可很好地与对应于感兴趣物体的较小梯度值联系起来。

10.5.2 水坝构建

在讨论之前，让我们考虑一下如何构建分水岭分割算法所需的水坝或分水线。水坝的构造是以二值图像为基础的，这种图像是二维整数空间 Z^2 的成员(见 2.4.2 节)。构建分离二元点集的水坝的最简方法是，使用形态学膨胀(见 9.2.2 节)。

图 10.55 说明了如何使用形态学膨胀来构建水坝的基础。图 10.55(a)显示了第 $n-1$ 步淹没的两个汇水盆地。图 10.55(b)显示了第 n 步淹没的结果。水已从一个盆地溢出到了另一个盆地，因此必须构建水坝来阻止这种情况的发生。为了与很快将要引入的符号一致，令 M_1 和 M_2 表示两个区域中的最小值点坐标的集合。然后，将汇水盆地中点的坐标集合与这两个在溢出的第 $n-1$ 步处的最小值点的坐标集合联系起来，在第 $n-1$ 步溢出处，两个最小值分别表示为 $C_{n-1}(M_1)$ 和 $C_{n-1}(M_2)$。它们是图 10.55(a)中的两个灰色区域。

令 $C[n-1]$ 表示这两个集合的并集。图 10.55(a)中有两个连通分量(关于连通分量的讨论，见 2.5.2 节)，而图 10.55(b)中只有一个连通分量，该连通分量包含了如虚线所示的前两个分量。两个连通分量变成一个连通分量的事实表明，两个汇水盆地在第 n 步淹没中已被聚合。令 q 表示这个连通分量。注意，来自 $n-1$ 步的两个连通分量可以使用简单的“与”操作($q \cap C[n-1]$)从 q 中提取。我们还注意到，属于单个汇水盆地的所有点形成了单个连通分量。

793 ~ 794

假设图10.55(a)中的每个连通分量被图10.55(c)中所示结构元膨胀满足两个条件：(1)膨胀必须受到 q 约束(这意味着在膨胀的过程中，结构元的中心只能位于 q 中的点处)，(2)不能对会引起正被膨胀集合聚合(变成单个连通分量)的那些点执行膨胀。图 10.55(d)表明，第一轮膨胀(显示为浅灰色)展开了每个原始连通分量的边界。注意，膨胀过程中，每个点都满足条件(1)，而在膨胀处理中，条件(2)未应用于任何点；这样，就均匀扩展了每个区域的边界。

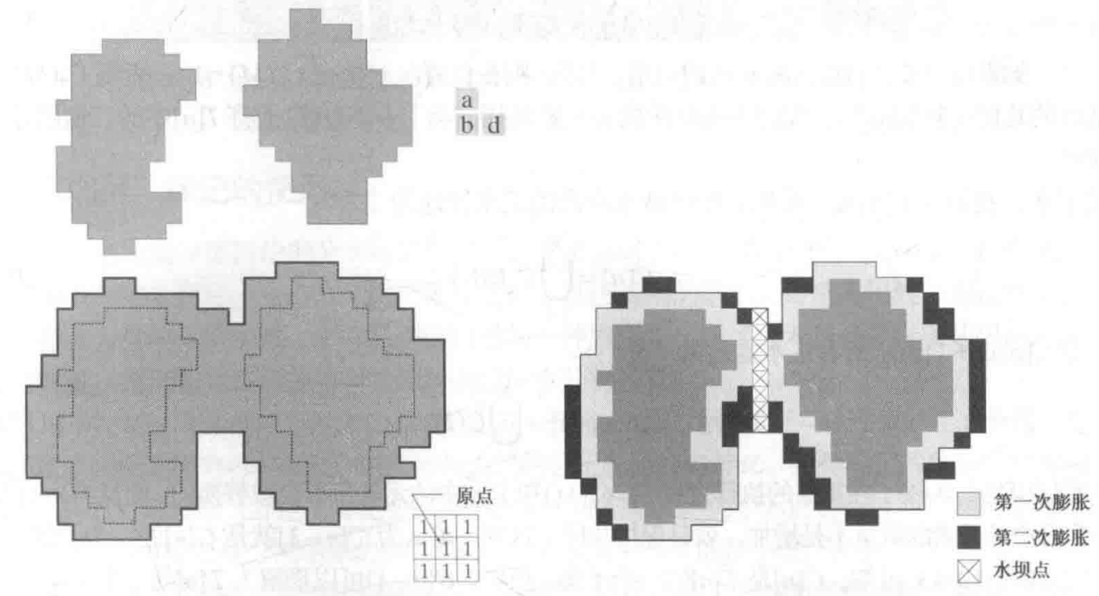


图 10.55　(a)第 $n-1$ 步淹没处，两个部分淹没的汇水盆地；(b)在淹没的第 n 步处，显示水已在两个盆地间溢出；(c)用于膨胀的结构元；(d)膨胀结果和水坝构建

在第二轮膨胀中(显示为黑色)，几个不符合条件(1)的点却符合条件(2)，这导致了图中所示的断开周界。很明显，q 中只有满足上述两个条件的点，才能描绘出图10.55(d)中由叉线所示的 1 像素宽的连通路径。在淹没的第 n 步，这条路径构成了所期望的分隔水坝。在这一淹没水平上的水坝构建，是在刚刚确定的这条路径上的所有点设置一个大于图像最大灰度值的值来完成的。所有水坝的高度通常设置为 1 加上图像中允许的最大灰度值。水位升高时，这样设置可以阻止洪水漫过所完成的水坝。注意，通过这一过程构建的水坝是连通分量很重要，它就是我们希望得到的分割边界。换句话说，这种方法消除了分割线断裂的问题。

尽管刚刚描述的过程只是基于一个简单的例子，但该方法同样适用于更为复杂的情形，包括使用图 10.55(c)中所示的 3×3 对称结构元。

10.5.3　分水岭分割算法

令 $M_1, M_2, \cdots, M_R$ 是图像 $g(x, y)$中区域最小值点的坐标集。如同在 10.5.1 节末尾说明的那样，这是一幅典型的梯度图像。令 $C(M_i)$是与区域最小值 M_i 相关的汇水盆地中的点的坐标集(回忆任何汇水盆地中的点都形成一个连通分量)。符号 min 和 max 表示 $g(x, y)$的最小值和最大值。最后，令 $T[n]$表示满足 $g(s, t) < n$ 的坐标(s, t)的集合，即

$$T[n] = \{(s, t) \mid g(s, t) < n\} \tag{10.5-1}$$

几何上，$T[n]$是 $g(x, y)$中位于平面 $g(x, y) = n$ 下方的点的坐标集。

随着水位以整数从 $n = \min + 1$ 到 $n = \max + 1$ 不断上升，地形将被水淹没。在淹没过程的任意步 n，算法都需要知道处在淹没深度之下的点的数量。从概念上讲，假设处在 $g(x, y) = n$ 平面之下的 $T[n]$

794 ~ 796

中的坐标被“标记”为黑色，所有其他坐标被标记为白色。然后，当我们以任何淹没增量 n 向下观察 xy 平面时，我们将会看到一幅二值图像，图像中的黑点对应于函数中平面 $g(x, y) = n$ 之下的点。这种解释对于理解下面的讨论很有帮助。

令 $C_n(M_i)$ 表示汇水盆地中与淹没阶段 n 的最小值 M_i 相关联的点的坐标集。参照前一段的讨论可知，$C_n(M_i)$ 可视为由下式给出的一幅二值图像：

$$C_n(M_i) = C(M_i) \cap T[n] \tag{10.5-2}$$

换句话说，如果$(x, y) \in C(M_i)$和$(x, y) \in T[n]$相“与”，则在位置(x, y)处有 $C_n(M_i) = 1$。，否则 $C_n(M_i) = 0$。这一结果的几何解释很简单，我们在淹没阶段 n 只需使用“与”(AND)算子将 $T[n]$中的二值图像分离出来即可。

接下来，我们令 $C[n]$表示阶段 n 中已被水淹没的汇水盆地的“并”：

$$C[n] = \bigcup_{i=1}^{R} C_n(M_i) \tag{10.5-3}$$

然后，令 $C[\max + 1]$表示所有汇水盆地的“并”：

$$C[\max+1] = \bigcup_{i=1}^{R} C(M_i) \tag{10.5-4}$$

可以证明(见习题 10.41)，在算法的执行过程中，$C_n(M_i)$和 $T[n]$中的元素是不会被替换的，而且当 n 增大时，这两个集合中的元素的数量不是增加，就是保持相同。这样，可认为 $C[n-1]$就是 $C[n]$的一个子集。根据式(10.5-2)和式(10.5-3)可知，$C[n]$是 $T[n]$的一个子集，所以，$C[n-1]$可以理解为 $T[n]$的一个子集。由此，我们得出一个重要的结果：$C[n-1]$中的每个连通分量都恰好包含在 $T[n]$的一个连通分量中。

寻找分水线的算法使用 $C[\min + 1] = T[\min + 1]$来初始化。然后，该算法进行递归处理，由 $C[n-1]$计算 $C[n]$。由 $C[n-1]$求得 $C[n]$的过程如下：令 Q 表示 $T[n]$中的连通分量的集合。然后，对于每个连通分量 $q \in Q[n]$，有如下三种可能性：

1. $q \cap C[n-1]$为空集。
2. $q \cap C[n-1]$包含 $C[n-1]$的一个连通分量。
3. $q \cap C[n-1]$包含 $C[n-1]$的一个以上的连通分量。

由 $C[n-1]$构建 $C[n]$取决于这三个条件中的哪个条件成立。遇到一个新的最小值时，条件 1 发生，这种情况下，连通分量 q 并入 $C[n-1]$中形成 $C[n]$。当 q 位于某些局部最小值的汇水盆地内时，条件 2 发生，这种情况下，q 并入 $C[n-1]$中形成 $C[n]$。当遇到全部或部分分隔两个或多个汇水盆地的山脊线时，条件 3 发生。进一步淹没会导致这些汇水盆地中的水位聚合。因此，必须在 q 内构筑一个水坝(如果涉及两个以上的汇水盆地，就要构筑多个水坝)以阻止汇水盆地间的水溢出。如前节中解释的那样，当我们使用元素为 1、大小为 3×3 的结构元来膨胀 $q \cap C[n-1]$且膨胀被限制到 q 时，可以构建一座 1 像素宽的水坝。

仅用 $g(x, y)$中对应于现有灰度值的 n 值就可改善算法的效率；根据 $g(x, y)$的直方图，我们可以确定这些值，以及最小值和最大值。

例 10.25　分水岭分割算法的说明。

分别考虑图 10.56(a)和(b)中的图像及其梯度。应用刚才讨论的分水岭算法得到了图 10.56(c)中的梯度图像的分水线(白色路径)。这些叠加到原图像上的分割边界示于图10.56(d)中。如本节开始时注释的那样，分割边界具有连通路径的重要特性。

796 ~ 798



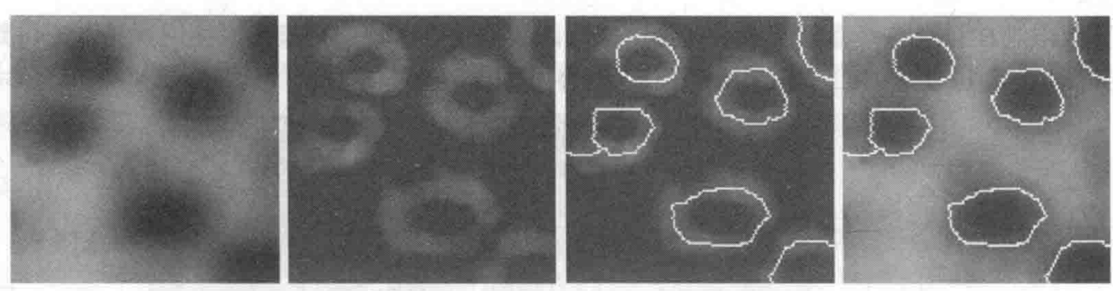


图10.56 (a)水滴的图像；(b)图像的梯度；(c)分水线；(d)叠加在原图像上的分水线（原图像由 CMM/Ecole des Mines de Paris 公司的 S. Beucher 博士提供）

10.5.4 标记的使用

直接应用前一节讨论的分水岭分割算法，通常会由于噪声和梯度的其他局部不规则性造成过度分割，如图 10.57 所示，过度分割的严重性足以令算法得到的结果变得毫无用处。在这种情况下，这意味着存在大量的分割区域。实际解决该问题的一种方案是，通过融入预处理步骤来限制允许存在的区域的数量，而预处理步骤是指将附加知识应用于分割过程。

用于控制过度分割的一种方法基于标记这一概念。标记是指属于一幅图像的连通分量。与感兴趣物体相联系的标记称为*内部标记*，与背景相关联的标记称为*外部标记*。选择标记的典型过程由两个主要步骤组成：(1)预处理；(2)定义标记必须满足的一个准则集合。为了对此进行说明，再次考虑图10.57(a)。导致图10.57(b)中过度分割结果的部分原因是，存在大量潜在的最小值。由于它们的尺寸，许多最小值是不相关的细节。就像先前的讨论中多次指出的那样，将很小的空间细节的影响降至最低的有效方法是，用一个平滑滤波器对图像进行过滤。在这种特殊情况下，这是一种合适的预处理方案。

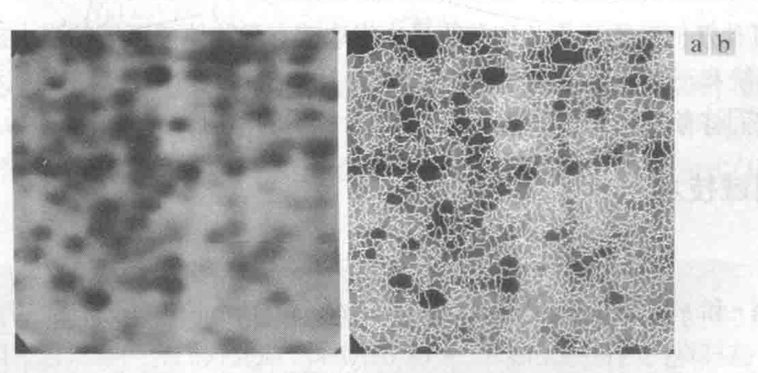


图 10.57 (a)电泳现象图像；(b)对梯度图像应用分水岭分割算法得到的结果。过度分割现象很明显（原图像由 CMM/Ecole des Mines de Paris 公司的 S. Beucher 博士提供）

假设我们将内部标记定义为：(1)被更高“海拔”点包围的区域；(2)区域中形成一个连通分量的那些点；(3)连通分量中所有的点有相同的灰度值。在图像经过平滑处理之后，由该定义导致的内部标记在图 10.58(a)中以浅灰色、斑点状区域显示。接着，在这些内部标记只能是在允许区域最小值的限制下，对平滑后的图像应用分水岭算法。图 10.58(a)显示了得到的分水线。将这些分水线定义为外部标记。请注意，沿分水线的点经过相邻标记间的最高点。

图 10.58(a)中的外部标记有效地将图像分割成了不同的区域，每个区域都包含一个内部标记和部分背景。这样，问题就简化为将每个区域划分为两部分：单个物体及其背景。我们可以根据这一简化过的问题，应用多种在本章先前讨论过的分割技术。另一种方法是，对各个区域简单地应用分水岭分割算法。换句话说，我们简单地求得平滑后图像的梯度［见如图 10.56(b)］，然后将

798 ~ 800

算法限制在只对在该特殊区域中包含这一标记的单一分水岭进行操作。使用这种方法得到的结果如图 10.58(b)所示。相对于图 10.57(b)，改善很明显。

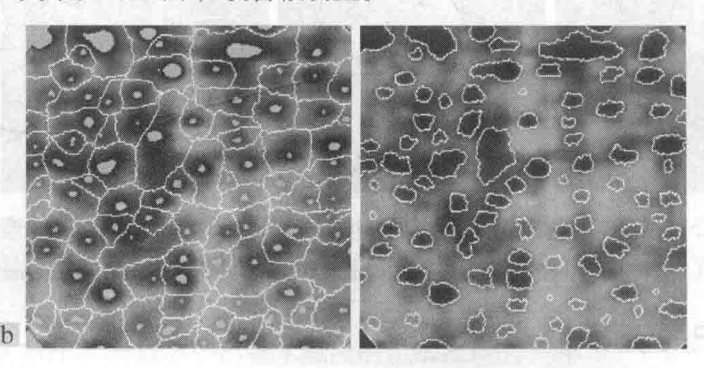


图 10.58 (a)显示有内部标记(浅灰色区域)和外部标记(分水线)的图像；(b)分割的结果。注意，注意其相对于图 10.47(b)的改进。(原图像由 CMM/Ecole des Mines de Paris 公司的 S. Beucher 博士提供)

标记的选择可以用基于灰度值和连通性的简单过程归类，如刚才所说明的那样，更复杂的描述涉及尺寸、形状、位置、相对距离、纹理内容等(见第11章中关于描绘子的内容)。关键是使用标记带来的与分割问题有关的先验知识。应该提醒读者注意的是：人类经常以先验知识用日常的视觉辅助进行各种分割和更高级的工作，最熟悉的例子之一就是上下文的使用。因此，分水岭分割方法提供了一种能有效使用这类知识的框架，这是这种方法的一个突出优点。

10.6 分割中运动的应用

运动是人类和其他任何动物从无关细节背景中提取感兴趣物体或区域的强大提示。在成像应用中，运动来自感觉系统和正被观看场景间的相对位移，比如机器人应用、自主导航和动态场景分析。在下面的几节中，我们将考虑运动在空间域和频率域分割中的运用。

10.6.1 空间域技术

基本方法

分别检测在时刻 t_i 和 t_j 获取的两帧图像 $f(x, y, t_i)$和 $f(x, y, t_j)$之间的变化的最简方法之一是，逐像素地比较这两幅图像。这样做的的目的是形成一幅差值图像。假设我们有一幅仅包含固定成分的参考图像。将这幅图像和一幅后续有着相同场景但包含一个运动物体的图像进行比较，去掉两幅图像中的固定元素，仅保留对应于非固定图像成分的非零项，就可得到两幅图像的差。

在时刻 t_i 和 t_j 获取的两幅图像间的差值图像定义为

$$d_{ij}(x, y)=\begin{cases}1, & |f(x, y, t_i)-f(x, y, t_j)|>T \\ 0, & \text{其他}\end{cases} \tag{10.6-1}$$

式中，T 是一个指定的阈值。注意，就像由该指定阈值确定的那样，仅当两幅图像间的灰度差在空间坐标(x, y)处明显不同时，$d_{ij}(x, y)$在该坐标处的值才为 1。假设所有图像的大小相同。最后，我们注意到，式(10.6-1)中坐标(x, y)的值跨越了这些图像的维度，所以差值图像 $d_{ij}(x, y)$与序列中其他图像的大小相同。

800 ~ 801

在动态图像处理中，$d_{ij}(x, y)$中值为 1 的所有像素被认为是物体运动的结果。这种方法只有在两幅图像都做了空间配准且在 T 值设定的范围内光照相对恒定的情况下才是可用的。实际上，$d_{ij}(x, y)$中的 1 值

项可能是由噪声造成的。通常，在差值图像中这些项是孤立的点，去除它们的一种简单方法是，在 $d_{ij}(x, y)$中构成由1组成的4连通或8连通区域，然后忽略1值项少于预定元素数量的任何区域。尽管这种方法会忽略掉小的和/或慢速运动的物体，但它改善了差值图像中剩余项实际是运动结果的机会。

累积差值

考虑一个图像帧序列 $f(x, y, t_1), f(x, y, t_2), \cdots, f(x, y, t_n)$，令 $f(x, y, t_1)$为参考图像。一幅累积差值图像(ADI)是通过将参考图像与序列中每幅后续图像进行比较形成的。当参考图像和序列中一幅图像之间在某个像素位置上出现一次差异时，累积图像中每个像素位置的计数器计数一次。因此，当第 k 帧图像与参考图像进行比较时，累积图像中一个给定像素的输入项，会给出在此位置上灰度与参考图像相应的像素值不同的次数[如由式(10.6-1)中的 T 确定的那样]。

考虑如下累积差值图像的三种类型：绝对 ADI、正 ADI 和负 ADI。假设运动物体的灰度值大于背景的灰度值，这三种 ADI 定义如下。令 $R(x, y)$表示参考图像，为简化符号，令 k 表示 t_k，则有 $f(x, y, k) = f(x, y, t_k)$。我们假设 $R(x, y) = f(x, y, 1)$。然后，对于任何 $k>1$，且记住 ADI 的值是计数，我们对(x, y)的所有相关值定义如下：

$$A_k(x, y) = \begin{cases} A_{k-1}(x, y) + 1, & |R(x, y) - f(x, y, k)| > T \\ A_{k-1}(x, y), & \text{其他} \end{cases} \tag{10.6-2}$$

$$P_k(x, y) = \begin{cases} P_{k-1}(x, y) + 1, & [R(x, y) - f(x, y, k)] > T \\ P_{k-1}(x, y), & \text{其他} \end{cases} \tag{10.6-3}$$

和

$$N_k(x, y) = \begin{cases} N_{k-1}(x, y) + 1, & [R(x, y) - f(x, y, k)] < -T \\ N_{k-1}(x, y), & \text{其他} \end{cases} \tag{10.6-4}$$

式中，$A_k(x, t)$, $P_k(x, y)$和 $N_k(x, y)$分别为遇到序列中的第 k 幅图像后的绝对 ADI、正 ADI 和负 ADI。

不难理解，这些 ADI 均从零值开始(计数)。还要注意这些 ADI 与序列中的图像大小相同。最后，我们注意到，如果背景像素的灰度值大于运动物体的灰度值，则式(10.6-3)和式(10.6-4)中不等式的顺序和阈值的符号是相反的。

例 10.26 绝对、正和负的累积差异图像的计算。

图 10.59 显示了以灰度图像形式显示的三种 ADI，图中矩形物体的大小为 75×50 像素，该物体以每帧 $5\sqrt{2}$ 个像素的速度向东南方向运动。图像的大小为 256×256 像素。我们注意到下列几点：(1) 正ADI 的非零区域等于运动物体的大小。(2) 正 ADI 的位置对应于参考帧中运动物体的位置。(3) 当运动物体与参考帧中的同一物体完全被移动时，正 ADI 中的计数停止增加。(4) 绝对 ADI 包含正 ADI 和负 ADI 的区域。(5) 运动物体的方向和速度可由绝对 ADI 和负 ADI 中的输入决定。

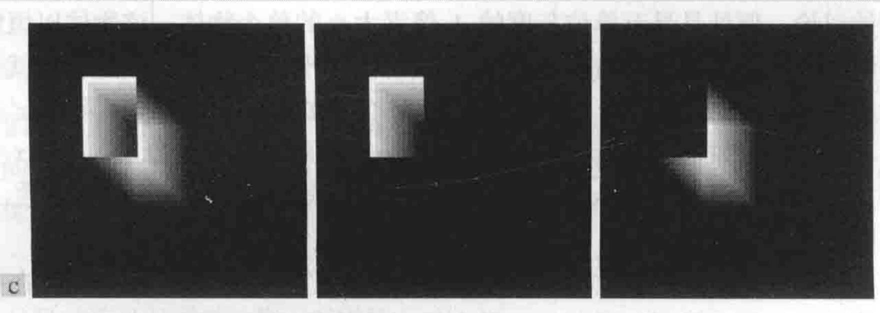


图 10.59 一个向东南方移动的矩形物体的 ADI：(a) 绝对 ADI；(b) 正 ADI；(c) 负 ADI

801 ~ 802

建立参考图像

前两节讨论的技术取得成功的关键是，找到一幅可以与后续图像进行比较的参考图像。在动态成像问题中，两幅图像的差值倾向于消除图像中的所有固定成分，而只留下对应于噪声和运动物体的图像元素。

实际上，仅用固定元素得到一幅参考图像并不总是可行的，所以有必要根据一组包含一个或多个运动物体的图像来构建一幅参考图像。这尤其适用于描述变化频繁的场景或需要频繁更新的场景的情况。生成参考图像的过程如下所述。考虑将图像序列中的第一幅图像作为参考图像。当一个非固定成分完全移出其参考帧的位置时，当前帧中对应的背景可复制到最初被参考帧中物体占据的位置。当所有的移动物体均已完全移出原来的位置时，就创建了一幅仅包含固定成分的参考图像。像前节中说明的那样，通过监视正 ADI 中的变化，可确定物体的位移。

例 10.27 构造一幅参考图像。

图 10.60(a)和(b)显示了十字路口的两幅图像帧。第一幅图像作为参考图像，第二幅图像描述了一段时间后的同一场景。为生成一幅静止图像，我们的目的是去掉参考图像中的主要运动物体。尽管存在其他较小的运动物体，但主要的运动特征是在十字路口从左到右运动的汽车。为了说明这一目的，我们将注意力放在这个物体上。通过监视正 ADI 的变化，如先前解释的那样，可以确定一个移动物体的初始位置。一旦确定了被这个物体占据的区域，使用减法可从图像中移除该物体。通过观察图像序列中正 ADI 停止变化的那一帧图像，我们可以从该幅图像中复制先前被初始帧中移动物体所占据的区域。然后，将该区域粘贴到被挖去物体的图像中，从而恢复该区域的背景。如果对所有运动的物体都执行这种操作，结果将是一幅只包含静止成分的参考图像，针对该参考图像，我们可在运动检测中与后续帧进行比较。在这种情况下，图 10.60(c)显示了去掉向东运动的汽车后得到的结果。

a b c





图 10.60 构造一幅静止参考图像：(a)和(b)序列中的两帧图像；(c)从图(a)中减去向东运动的汽车并使用图(b)中的对应区域恢复背景后的结果(原图像由 Jain 和 Jain 提供)

10.6.2 频率域技术

本节考虑通过傅里叶变换公式来确定运动的问题。考虑一个序列 $f(x, y, t)$, $t = 1, 2, \cdots, K-1$，该序列是由一台固定照相机拍摄的大小为 $M \times N$ 的 K 帧数字图像。我们从假设所有帧都有零灰度的均匀背景开始讨论。例外是具有单位灰度的 1 像素大小的单个物体，该物体以恒定速度运动。假设对于第一帧($t = 0$)，物体在位置(x', y')处，并且将图像平面投影到 x 轴上；也就是说，对图像中的一列像素的灰度相加。这一操作产生一个带有 M 个零值项的一维阵列，除了 x' 处，它是单点物体的 x 坐标。现在，如果我们用 $\exp[\mathrm{j}2\pi a_1 x\Delta t]$ 乘以一维阵列中的每个分量，其中 $x = 1, 2, \cdots, M-1$，并把结果相加，我们会得到单项 $\exp[\mathrm{j}2\pi a_1 x'\Delta t]$。在这种表示中，$a_1$ 是一个正整数，Δt 是不同帧之间的时间间隔。

802~804

假设在第二帧中（$t = 1$），物体已运动到坐标 $(x'+1, y')$ 处；也就是说，物体在平行于 x 轴的方向移动了 1 像素的位置。然后，重复上一段中提到的投影过程，得到和 $\exp[\mathrm{j}2\pi a_1 (x'+1)\Delta t]$。若物

体每帧继续移动 1 像素，则在任何整数瞬间 t，结果为 $\exp[\mathrm{j}2\pi a_1(x'+t)\Delta t]$，使用欧拉公式，可将它表示为

$$e^{\mathrm{j}2\pi a_1(x'+t)\Delta t}=\cos[2\pi a_1(x'+t)\Delta t]+\mathrm{j}\sin[2\pi a_1(x'+t)\Delta t] \tag{10.6-5}$$

式中，$t=0,1,\cdots,K-1$。换句话说，这一过程得到一条频率为 a_1 的复正弦曲线。如果物体在帧间移动 V_1 个像素(在 x 方向)，那么正弦曲线的频率为 V_1a_1。因为 t 在 0 和 $K-1$ 之间以整数增量变化，限制 a_1 只取整数值会使复正弦曲线的离散傅里叶变换具有两个波峰——一个位于频率 V_1a_1 处，另一个位于 $K-V_1a_1$ 处。如 4.6.4 节中讨论的那样，后一个波峰是由离散傅里叶变换的对称性产生的，故可以忽略。因此，在傅里叶频谱中找到一个波峰，就得到 V_1a_1。用 a_1 除该量得到 V_1，它就是 x 方向的速度分量，这里假设帧速率是已知的。采用类似的方法可以得到 y 方向的速度分量 V_2。

没有运动发生的帧序列会产生相同的指数项，其傅里叶变换在频率 0(单个直流项)处有一个波峰。所以，由于迄今为止我们讨论的都是线性操作，故在任意静止背景中涉及一个或多个运动物体的一般情况下，对应于静止图像分量的直流处，傅里叶变换有一个波峰，且这些位置的波峰与物体的速度成正比。

上述概念可以总结如下：对于由 K 幅大小为 $M\times N$ 的数字图像组成的序列，在任何整数的瞬时点上，图像投影到 x 轴上的加权和为

$$g_x(t,a_1)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y,t)e^{\mathrm{j}2\pi a_1 x\Delta t}\quad t=0,1,\cdots,K-1 \tag{10.6-6}$$

类似地，y 轴上的投影和为

$$g_y(t,a_2)=\sum_{y=0}^{N-1}\sum_{x=0}^{M-1}f(x,y,t)e^{\mathrm{j}2\pi a_2 y\Delta t}\quad t=0,1,\cdots,K-1 \tag{10.6-7}$$

其中，如已经解释过的那样，a_1 和 a_2 都是正整数。

式(10.6-6)和式(10.6-7)的一维傅里叶变换分别为

$$G_x(u_1,a_1)=\sum_{t=0}^{K-1}g_x(t,a_1)e^{-\mathrm{j}2\pi u_1 t/K}\quad u_1=0,1,\cdots,K-1 \tag{10.6-8}$$

和

$$G_y(u_2,a_2)=\sum_{t=0}^{K-1}g_y(t,a_2)e^{-\mathrm{j}2\pi u_2 t/K}\quad u_2=0,1,\cdots,K-1 \tag{10.6-9}$$

实际上，如 4.11 节中讨论的那样，这些变换的计算是使用 FFT 算法执行的。

频率和速度的关系是

$$u_1=a_1V_1 \tag{10.6-10}$$

和

$$u_2=a_2V_2 \tag{10.6-11}$$

在该式中，速度的单位是像素数/总帧时间。例如，$V_1=10$ 可以解释为在 K 帧中物体运动了 10 个像素。对于匀速拍摄的图像帧，实际的物理速度取决于帧速率和像素之间的距离。因此，若 $V_1=10$，$K=30$，帧率是两幅图像/秒，像素之间的距离是 0.5 m，则 x 方向的实际物理速度是

$$V_1=(10\text{ 个像素})(0.5\text{ m/像素})(2\text{ 帧/s})/(30\text{ 帧})=1/3\text{ m/s}$$

速度的 x 分量的符号可通过计算下式得到：

804~805

$$S_{1x} = \left.\frac{\mathrm{d}^2\,\mathrm{Re}[g_x(t,\,a_1)]}{\mathrm{d}t^2}\right|_{t=n} \tag{10.6-12}$$

和

$$S_{2x} = \left.\frac{\mathrm{d}^2\,\mathrm{Im}[g_x(t,\,a_1)]}{\mathrm{d}t^2}\right|_{t=n} \tag{10.6-13}$$

因为 g_x 是正弦函数，故可以证明(见习题 10.47)，若速度分量 V_1 为正，则在时间 n，S_{1x} 和 S_{2x} 在任意点将有相同的符号。反之，S_{1x} 和 S_{2x} 的符号相反则表明速度分量 V_1 为负。如果 S_{1x} 或 S_{2x} 为零，那么我们考虑下一个最近的时间点 $t = n \pm \Delta t$。类似的说明适用于计算 V_2 的符号。

例 10.28　通过频率域检测较小的运动物体。

图 10.61 到图 10.64 说明了刚才推导方法的有效性。图10.61 显示了一组 32 帧 LANDSAT 图像序列中的一幅图像，该图像是通过将白噪声添加到一幅参考图像中产生的。图像序列中包含一个叠加的运动物体，物体在 x 方向以每帧 0.5 个像素的速度运动，在 y 方向以每帧 1 个像素的速度运动。在图 10.62 中用圆圈标出的物体，在展开的较小(9 像素)区域内呈高斯灰度分布，该分布用肉眼不易辨别。图 10.63 和图 10.64 显示了分别令 $a_1 = 6$ 和 $a_2 = 4$ 计算式(10.6-8)和式(10.6-9)时得到的结果。根据图 10.63 中 $u_1 = 3$ 处的波峰，由式(10.6-10)得到 $V_1 = 0.5$。同样，根据图 10.64 中 $u_2 = 4$ 处的波峰，式(10.6-11)得到 $V_2 = 1.0$。

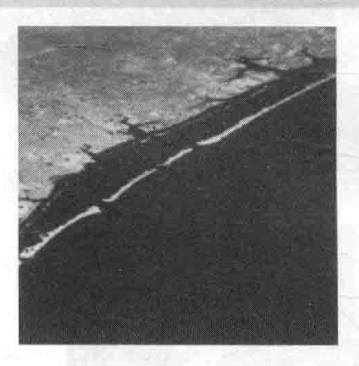


图 10.61　LANDSAT 的图像帧(原图像由 Cowart, Snyder 和 Ruedger 提供)

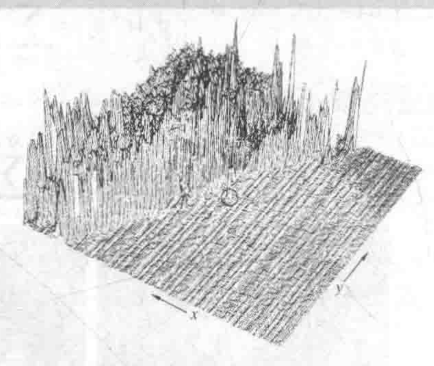


图 10.62　图 10.61 中图像的灰度图，目标已用圆圈标出(原图像由 Rajala, Riddle 和 Snyder 提供)

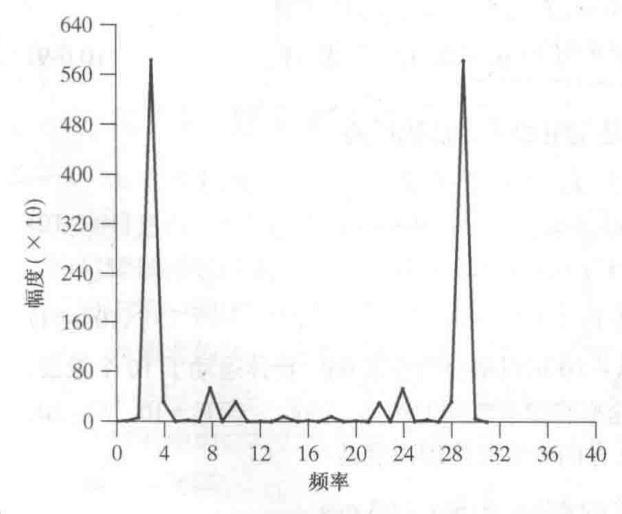




图 10.63　式(10.6-8)的谱，在 $u_1 = 3$ 处显示了一个波峰(原图像由 Rajala, Riddle 和 Snyder 提供)

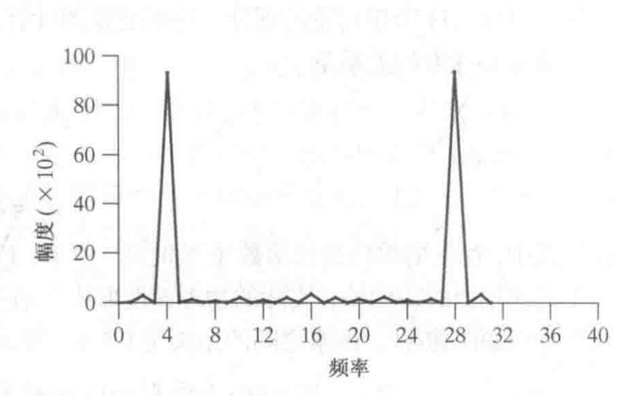


图 10.64　式(10.6-9)的谱，在 $u_2 = 4$ 处显示了一个波峰(原图像由 Rajala, Riddle 和 Snyder 提供)

借助于图 10.63 和图 10.64，我们可以解释选择 a_1 和 a_2 的原则。例如，假设我们用 $a_2 = 15$ 取代 $a_2 = 4$。在这种情况下，图 10.64 中的波峰将位于 $u_2 = 15$ 和 17 处，因为 $V_2 = 1.0$，这将是一种严重的混淆结果。如 2.5.4 节讨论的那样，混淆是由欠采样造成的（当 u 的范围由 K 决定时，当前讨论中，图像帧数太少）。因为 $u = aV$，故一种可能是将 a 选择为最接近 $a = u_{max}/V_{max}$ 的整数，其中 u_{max} 是由 K 建立的混淆频率限制，而 V_{max} 是期望的物体的最大速度。

小结

在大多数自动图像模式识别和场景分析应用中，图像分割是一种基本的预处理步骤。如前面各节中各个例子说明的那样，选择何种分割技术主要取决于所考虑问题的特殊性。本章中讨论的各种方法尽管并不十分详尽，但却是在实际应用中普遍使用的代表性技术。下面的参考文献可作为读者进一步研究该主题的基础。

参考文献

由于分割在自动图像处理领域中所起的核心作用，大多数关于图像处理、图像分析和计算机视觉的著述中都涵盖了这一主题。以下图书为这一主题提供了补充和/或辅助阅读材料：Umbaugh[2005]、Davies[2005]、Gonzalez, Woods and Eddins[2004]、Shapiro and Stockman[2001]、Sonka et al.[1999]、Petrou and Bosdogianni[1999]。

使用模板检测灰度的不连续性（见 10.2 节）已有较长的历史。这些年来，人们提出了无数模板：Roberts[1965]、Prewitt[1970]、Kirsh[1971]、Robinson[1976]、Frei and Chen[1977]和 Canny[1986]。Fram and Deutsch[1975]中包含了大量的模板及这些模板的性能评价。模板性能存在的问题，特别是对边缘检测来说，仍是相当重要的研究领域，详见 Qian and Huang[1996]、Wang et al.[1996]、Heath et al.[1997, 1998]和 Ando[2000]。彩色图像边缘检测在许多感知的应用中正逐渐普及，详见 Salians, Abidi and Gonzalez[1996]、Zugaj and Lattuati[1998]、Mirmehdi and Petrou[2000]和 Plataniotis and Venetsanopoulos[2000]。图像特征和模板性能之间的相互影响也是当前人们感兴趣的主题，见 Ziou[2001]。书中关于拉普拉斯算子的零交叉性质的介绍，源于 Marr and Hildredth[1980]和 Marr[1982]，也可参见 Clark[1989]关于使用零交叉算法进行边缘鉴别的论文（Piech[1990]对 Clark 的论文给出了部分修正）。如 10.2 节中提到的那样，通过高斯拉普拉斯的零交叉方法是一种重要的方法，其相对性能仍是一个活跃的研究主题（Gunn[1998, 1999]）。如其名称暗示的那样，在 10.2.6.节讨论的坎尼检测器基于坎尼的论文[1986]。这一主题 20 年后的例子见 Zhang and Rockett[2006]。

霍夫变换（Hough[1962]）是进行全局像素连接和曲线检测的一种实用方法。本章讨论的这种基本变换的大量推广已提出多年。例如，Lo and Tsai[1995]讨论了检测细线的一种方法，Guil et al.[1995, 1997]解决了霍夫变换的快速实现和简单曲线的检测问题，Daul et al.[1998]将霍夫变换进一步推广到了检测椭圆形弧线，Shapiro[1996]则解决了对灰度图像实现霍夫变换的问题。

如 10.3 节开始时提到的那样，阈值处理技术享有重要的普及地位，因为这种技术实现简单。关于这一主题，许多文献中报告了大量的研究成果。在 Sahoo et al.[1988]和 Lee et al.[1990]的综述性文章中，可以看到对这类文献的较好评价。除了本章中讨论的技术外，处理光照和反射影响（10.3.1 节）所使用的其他方法，见 Perez and Gonzalez[1987]、Parker[1991]、Murase and Nayar[1994]、Bischsel[1998]、Drew et al.[1999]和 Toro and Funt[2007]。关于 10.3.2 节中的其他读物，见 Jain et al.[1995]。

最优全局阈值处理（10.3.3 节）的早期工作已在 Chow and Kaneko[1972]中举例说明（我们将在 12.2.2 节讨论这种方法）。虽然理论上它是最佳的，但这种方法在灰度阈值处理中的应用却受到了限制，因为需要估计概率密度函数。10.3.3 节中源自 Otsu[1979]的最佳方法得到了人们的广泛接受，因为它使优秀的性能和实现

806 ~ 808

简单结合起来,且仅要求估计图像的直方图。使用预处理(10.3.4 节和 10.3.5 节)的基本概念可回溯至 White and Rohrer[1983]的早期论文，该论文在解决困难的分割问题中，综合使用了阈值处理、梯度和拉普拉斯算子。前面三篇文章中讨论的方法，与 Cheriet et al.[1998]、Sauvola and Pietikainen[2000]、Liang et al.[2000]和 Chan et al.[2000]在阈值处理方面所做的工作几乎相差了近 20 年，因此在图像分割能力方面比较这些方法的基本相似性是很有趣的。关于多阈值处理(见 10.3.6 节)的其他读物，见 Yin and Chen[1997]、Liao et al.[2001]和 Zahara et al.[2005]。关于可变阈值处理(见 10.3.7 节)的其他读物，见 Parker[1997]和 Delon et al.[2007]。

面向区域分割这一主题的早期著述见 Fu and Mui[1981]。为分割目的而综合区域和边界信息的早期研究见 Haddon and Boyce[1990]和 Pavlids and Liow[1990]。Hojjatoleslami and Kittler[1998]提出的一种较新区域生长方法也很重要。近期基本涵盖面向区域分割概念的文献，见 Shapiro and Stockman[2001]和 Sonka et al.[1999]。

10.5 节所示的分水岭分割是一个强大的概念。涉及分水岭分割的早期文献见 Serra[1988]、Beucher[1990]和 Beucher and Meyer[1992]。Baccar et al.[1996]讨论了基于数据融合和形态学分水岭的分割方法。完全致力于分水岭分割的专刊 *Pattern Recognition*[2000]表明，10 年来该领域的研究进展相当明显。如我们在 10.5 节中讨论的那样，使用分水岭方法的关键问题之一是过度分割问题。Najmanand and Schmitt[1996]、Haris et al.[1998]和 Bleau and Leon[2000]是处理这种问题的例证。Bieniek and Moga[2000]讨论了一种基于连通分量的分水岭分割算法。

10.6.1 节中的内容来自 Jain, R.[1981]，也可参阅 Jain, Kasturi and Schunck[1995]。10.6.2 节中的内容来自 Rajala, Riddle and Snyder[1983]，也可参阅 Shariat and Price[1990]和Cumani et al.[1991]。关于运动估计的其他读物，见 Sonka et al.[1999]、Shapiro and Stockman[2001]、Snyder and Qi[2004]和 Davies [2005]，也可参阅 Alexiadis and Sergiadis[2007]。

习题

★10.1 证明式(10.2-2)的正确性。(提示：用泰勒级数展开，并仅保留线性项。)

★10.2 设有一幅包含有水平直线、垂直直线、45°方向直线和 – 45°方向直线的二值图像。给出一组大小为 5×5 的模板，要求这些模板可用于检测这些直线中的 1 像素间断。假设直线和背景的灰度分别是 1 和 0。

10.3 提出一种检测二值图像的线段中长度从 1 到 K 个像素的裂缝的技术。假设线宽为 1 像素。你的技术要以 8 邻域连通性分析为基础，而不能试图构造检测裂缝的模板。

10.4 参考图 10.7，回答下列问题。

★(a) 图 10.7(e)中某些连接焊点和中心元素的线是单线，而其他线是双线。解释一下为什么。

(b) 提出一种消除图 10.7(f) 中那些 – 45°方向线的部分分量的方法。

10.5 参考图 10.8 中的边缘模型。

★(a) 假设我们使用图 10.14 中的 Prewitt 算子来计算每个模型的梯度幅度。请画出通过梯度图像中心的水平剖面线。

(b) 画出每个对应角度图像的水平剖面。

(注意：在回答该问题时不要生成梯度图像和角度图像。简单地提供剖面线的图形，显示像你预期的幅度和角度的剖面线即可。)

808 ~ 809

10.6 考虑穿过一幅二值图像中部的一条水平灰度剖面线，二值图像中包含一个垂直通过图像中心的台阶边缘。画出已被一个大小为 $n\times n$ 且系数等于 $1/(n^2)$ 的平均模板模糊后的图像的剖面线。为简单

起见，假设图像已被标定，即其灰度级在边缘的左侧为 0，在右侧为 1。此外，假设模板的尺寸远小于图像的尺寸，以便在靠近水平灰度剖面的中心处不必关注图像的边界效应。

★10.7 假设我们已使用示于右图中的边缘模型代替了图10.10 中的斜坡模型。请写出每个剖面的梯度和拉普拉斯算子。

图像

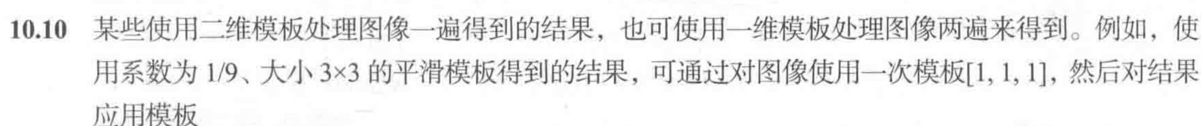
一条水平线的剖面

10.8 参照图 10.14，回答下列问题。

(a) 假设使用 Sobel 模板得到了 g_x 和 g_y。证明在这种情况下，使用式(10.2-10)和式(10.2-20)计算得到的梯度幅度可得出相同的结果。

(b) 证明对于 Prewitt 模板也是正确的。

★10.9 证明图 10.14 和图 10.15 中的 Sobel 模板和 Prewitt 模板仅对水平边缘、垂直边缘和 ±45°方向边缘给出各向同性的结果。



10.10 某些使用二维模板处理图像一遍得到的结果，也可使用一维模板处理图像两遍来得到。例如，使用系数为 1/9、大小 3×3 的平滑模板得到的结果，可通过对图像使用一次模板[1, 1, 1]，然后对结果应用模板



$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

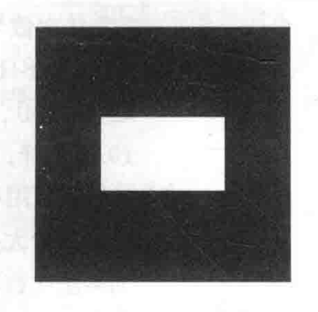
最终结果再乘以 1/9 得到。证明 Prewitt 模板(见图 10.14)的响应可通过使用一次形如[−1, 0, 1]的差分模板(或其对应的正交阵)，然后使用一次形如[1, 2, 1]的平滑处理模板(或其对应的正交阵)得到。

10.11 3×3 罗盘梯度算子设计用来度量 8 个方向上的梯度：E, NE, N, NW, W, SW, S 和 SE。

★(a) 使用系数值 0, 1, −1, −2 或 2，给出这 8 个算子的形式。

(b) 指定每个模板的梯度向量方向，记住梯度方向与边缘方向是正交的。

10.12 右图所示二值图像中，矩形的大小为 $m \times n$ 像素。



(a) 以式(10.2-20)给出的近似为基础，这幅图像的梯度幅度像什么？假设 g_x 和 g_y 是使用 Prewitt 算子得到的。说明梯度图像中所有相应的不同像素值。

(b) 画出用式(10.2-11)计算得到的边缘方向的直方图。精确标出直方图中每个分量的高度。

(c) 根据式(10.2-7)给出的近似，该图像的拉普拉斯图像是什么？说明拉普拉斯图像中所有相应的不同像素值。

10.13 假设图像 $f(x, y)$与大小为 $n \times n$ 的模板(系数为 $1/n^2$)卷积后，产生了一幅平滑后的图像 $\bar{f}(x, y)$。

★(a) 以模板尺寸的函数的形式，推导平滑后图像的边缘强度(边缘幅度)的表达式。为简单起见，假设 n 是奇数，且边缘是使用如下偏导数得到的：

$$\partial\bar{f}/\partial x = \bar{f}(x+1, y) - \bar{f}(x, y) \quad 和 \quad \partial\bar{f}/\partial y = \bar{f}(x, y+1) - \bar{f}(x, y)$$

(b) 证明平滑后的图像的最大边缘强度与原图像的最大边缘强度之比为 $1/n$。换句话说，边缘强度与平滑模板的尺寸成反比。

10.14 参考式(10.2-23)：

★(a) 证明一个高斯算子的拉普拉斯 $\nabla^2 G(x, y)$为零。

809 ~ 811

(b) 证明用这个算子卷积的任何图像的平均值也为零。(提示：考虑在频率域中解决该问题，使用卷积定理，以及一个函数的平均值与其在原点处计算的傅里叶变换成正比这一事实。)

(c) 一般情况下，如果我们(1)使用图 10.4(a)中的模板来计算使用大小为 3×3 的拉普拉斯模板的高斯低通滤波器的拉普拉斯，并(2)将得到的结果与任何图像卷积，那么请问(b)还正确吗？请解释。(提示：参见习题 3.16。)

10.15 参考图 10.22(c)。

(a) 解释为什么边缘构成闭合的轮廓线。

★(b) 寻找边缘位置的零交叉方法总会导致闭合轮廓吗？请解释。

10.16 我们时常发现文献中推导高斯拉普拉斯(LoG)时，都用如下表达式开始：

$$G(r) = e^{-r^2/2\sigma^2}$$

式中 $r^2 = x^2 + y^2$。然后，用二阶偏导数得到 LoG：$\nabla^2 G(r) = \partial^2 G/\partial r^2$。最后，使用 $x^2 + y^2$ 替代 r^2 得到了(不正确的)结果：

$$\nabla^2 G(x, y) = [(x^2 + y^2 - \sigma^2)/\sigma^4]\exp[-(x^2 + y^2)/2\sigma^2]$$

请推导该结果并解释该表达式与式(10.2-23)不同的原因。

10.17 (a) 推导式(10.2-27)。

(b) 令 $k = \sigma_1/\sigma_2$ 表示 DoG 函数的标准差比率。请使用 k 和 σ_2 来表示式(10.2-27)。

10.18 在下文中，假设 G 和 f 分别是大小为 $n\times n$ 和 $M\times N$ 的离散阵列。

★(a) 证明式(10.2-21)中的高斯函数与图像 $f(x, y)$ 的二维卷积，可以表示为先沿 $f(x, y)$ 的行(列)进行一维卷积，然后沿所得结果的列(行)进行一维卷积(关于离散卷积的讨论，见 3.4.2 节)。

(b) 推导出使用(a)中的一维卷积方法相对于直接执行二维卷积的优势的表达式。假设 $G(x, y)$ 被取样为一个大小为 $n\times n$ 的阵列，而 $f(x, y)$ 是一个大小为 $M\times N$ 的阵列。计算优势是二维卷积要求的乘法次数与一维卷积要求的乘法次数之比。

10.19 (a) 证明 Marr-Hildreth 算法的步骤 1 和步骤 2 可以通过使用 4 次一维卷积来实现。(提示：参考习题10.18(a)，将拉普拉斯算子表示为由式(10.2-5)和式(10.2-6)给出的两个偏导数之和，如习题 10.10 那样，并使用一个一维模板实现每个导数。)

(b) 推导出使用(a)中的一维卷积方法相对于直接执行二维卷积的优势的表达式。假设 $G(x, y)$ 被取样为一个大小为 $n\times n$ 的阵列，而 $f(x, y)$ 是一个大小为 $M\times N$ 的阵列。计算优势是二维卷积要求的乘法次数与一维卷积要求的乘法次数之比(见习题 10.18)。

10.20 (a) 用一维卷积代替二维卷积，用公式表示坎尼算法的步骤 1 和步骤 2 中梯度幅度图像的计算。

(b) 相对于使用二维卷积实现，使用一维卷积方法的计算优势是什么？假设步骤 1 中的二维高斯滤波器被取样为一个大小为 $n\times n$ 的阵列，且输入图像的大小为 $M\times N$。将计算优势表示为每种方法所要求的乘法次数之比。

10.21 参考图 10.8 中的三个垂直边缘模型和对应的剖面。

★(a) 假设我们使用 Sobel 模板来计算三个边缘模型的梯度幅度。画出三幅梯度图像的水平灰度剖面线。

★(b) 画出三幅拉普拉斯图像的水平灰度剖面线，假设拉普拉斯是使用图 10.4(a)中的 3×3 模板计算的。

811 ~ 812

★(c) 仅使用 Marr-Hildreth 边缘检测器的前两步产生的图像，重复上一问题。

(d) 对于坎尼边缘检测器的前两步重复上述问题。你可以忽略角度图像。

(e) 针对坎尼边缘检测器，画出角度图像的水平剖面线。(注意：回答这一问题时不必生成图像，只须简单地画出剖面线，并说明你所期望的图像剖面线像什么。)

10.22 参考 10.2.7 节中讨论的霍夫变换。

(a) 从 $y = ax + b$ 的斜截式形式，导出一条线的法线表示的一般过程。

★(b) 求直线 $y = -3x + 2$ 的法线表示。

★**10.23** 参考 10.2.7 节中讨论的霍夫变换。

(a) 解释为什么图 10.33(a)中的点 1 的霍夫映射在图 10.33(b)中为一条直线。

(b) 这是仅能产生这样结果的点吗？请解释。

(c) 以图 10.33(b)中标为 Q 的曲线为例，解释反射的邻接关系。

10.24 证明实现 10.2.7 节所述累加单元方法所需的操作次数与图像平面(即 xy 平面)中的背景点数量 n 呈线性关系。

10.25 图像分割技术的重要应用领域是处理由所谓的泡沫室活动导致的图像。这些图像是在高能物理实验中得到的，在这种实验中，一束已知粒子直接打到一个已知核子上。一次典型的活动由任何一次碰撞活动中的输入轨迹以及粒子由撞击点分散发出的二次轨迹组成。请提出一种分割方法来检测所有的轨迹，这些轨迹至少包含 100 个像素并与水平方向有如下 6 个方向的夹角：±20°，±40° 和±60°。在这 6 个方向上，允许的估计误差为±5°。一条有效的轨迹必须至少要有 100 个像素长，间隙不多于 3 个，并且每个间隙不超过 10 个像素的长度。你可以假设图像已进行了预处理，形成了二值图像，并且除了发出轨迹的撞击点以外，所有轨迹均为 1 像素宽。你的过程应能分辨原点不同但方向相同的轨迹。(提示：以霍夫变换作为你的解决方案的基础。)

★**10.26** 重新说明 10.3.2 节中的全局阈值处理算法，以便它使用一幅图像的直方图来代替该图像本身。

★**10.27** 证明10.3.2 节中的基本全局阈值处理算法在有限几步内收敛。(提示：使用来自习题 10.26 的直方图表达式。)

10.28 给出一种解释，为什么 10.3.2 节的基本全局阈值处理算法的初始阈值必须在图像的最小值和最大值之间。(提示：构建一个例子来说明阈值选在该范围之外算法会失败。)

★**10.29** 使用 10.3.2 节的基本全局阈值处理算法得到的阈值与起始点无关吗？如果你的回答是“是”，证明它，如果回答是“不”，给出一个例子。

10.30 你可以假设下列两种情况，在迭代中阈值限制在开区间 $(0, L-1)$ 内。

★(a) 如果图像的直方图在所有可能的灰度级上都是均匀的，请证明 10.3.2 节的基本全局阈值处理算法将收敛到图像的平均灰度 $(L-1)/2$ 处。

(b) 如果一幅图像的直方图是双模式的，且具有关于其均值对称的相同模式，请证明基本全局阈值处理算法将收敛到这两个模式的均值之间的中点。

10.31 参考 10.3.2 节中的阈值处理算法。假设在给定的问题中，直方图是双模式的，且这两个模式是形如 $A_1\exp[-(z-m_1)^2/2\sigma_1^2]$ 和 $A_2\exp[-(z-m_2)^2/2\sigma_2^2]$ 的高斯曲线。假设 $m_1 > m_2$，且初始的 T 在图像的最大灰度值和最小灰度值之间。当算法收敛时，给出以下陈述为真的条件(用这些曲线的参数表示)：

(a) 阈值等于 $(m_1 + m_2)/2$。

(b) 阈值在 m_2 的左边。

(c) 阈值位于区间 $(m_1 + m_2)/2 < T < m_1$ 内。

如果这些条件的任意一条不存在，请说明原因。

★**10.32** (a) 证明式(10.3-15)的第一行来自式(10.3-14)、式(10.3-10)和式(10.3-11)。

(b) 证明式(10.3-15)的第二行来自第一行。

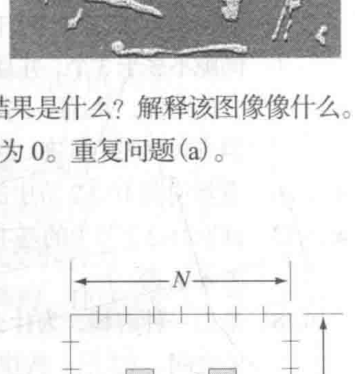
812 ~ 814

10.33 证明对于区间 $0 \leqslant k \leqslant L-1$ 内的 k 值，式(10.3-18)总存在一个最大值。

10.34 参考式(10.3-20)，预设一个参量，可确定 $0 \leqslant \eta(k) \leqslant 1$，$k$ 在 $0 \leqslant k \leqslant L-1$ 范围内，仅在图像具有恒定灰度时可达到最小值，并且仅在图像有两个值 0 和 $L-1$ 时最大值才发生。

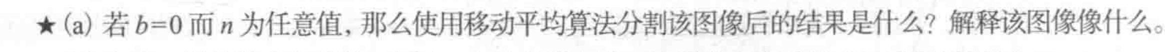
★10.35 (a) 假设一幅图像 $f(x, y)$ 的灰度在区间[0, 1]内，并且阈值 T 成功地将该图像分割为物体和背景。证明阈值 $T' = 1-T$ 将成功地把 $f(x, y)$ 的负图像分割为相同的区域。这里，“负”这一术语被用于 3.2.1 节定义的场景。

(b) 在(a)中，把一幅图像映射为其负图像的灰度变换函数，是一个负斜向的线性函数。说明一个任意的灰度变换函数必须满足关于将被保留的阈值 T 对原始图像可分割性的条件。灰度变换后的阈值将是什么值？



10.36 右图所示图像中的物体和背景，在标度范围[0, 255]内具有的平均灰度分别为 180 和 70。该图像被均值为 0、标准差为 10 个灰度级的高斯噪声污染。请提出一种正确分割率为 90%或更高百分比的阈值处理方法。(回忆一下，高斯曲线下 99.7%的面积位于均值的 $\pm 3\sigma$ 区间内，其中 σ 是标准差。)

10.37 参考图 10.37(b)中的灰度斜坡图像和 10.3.7 节中讨论的移动平均算法。假设图像的尺寸为 400×700 像素，其最小值和最大值分别为 0 和 1，这里的 0 项仅包含在第一列中。



★(a) 若 $b=0$ 而 n 为任意值，那么使用移动平均算法分割该图像后的结果是什么？解释该图像像什么。

(b) 现在，斜坡的方向反转，因此，最左侧的值为 1 而最右侧的值为 0。重复问题(a)。

(c) 重复问题(a)，但 $n=4, b=1$。

(d) 重复问题(a)，但 $n=100$，$b=1$。

10.38 提出一个区域生长算法来分割习题 10.36 中的图像。

★10.39 使用10.4.2节中讨论过的分裂和聚合过程来分割右图所示的图像。如果 R_i 中的所有像素都有相同的灰度，则令 $Q(R_i) = \text{TRUE}$。给出对应于你的分割的四叉树。

10.40 考虑例 10.24 中对天鹅星座环图像的稀疏区域进行分割得到的由 1 构成的区域。提出一种将该区域作为模板来隔离图像中的三个主要成分的技术：(1)背景，(2)致密内部区域，(3)稀疏外部区域。

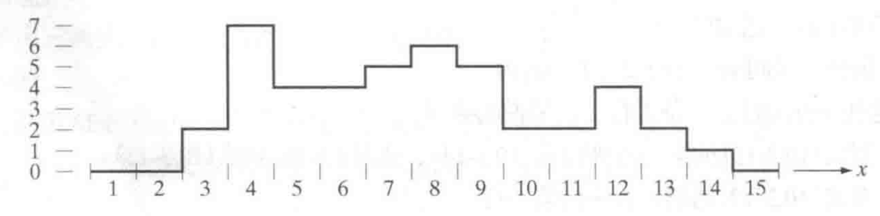
10.41 参考 10.5.3 节中的讨论。

★(a) 说明在分水岭算法执行的过程中，$C_n(M_i)$ 和 $T[n]$ 的元素从不被替换。

(b) 说明当 n 增加时，集合 $C_n(M_i)$ 和集合 $T[n]$ 中的元素数量随着 n 的增加，不是增加，就是保持不变。

10.42 在 10.5 节中，使用分水岭分割算法得到的边界形成了一个闭环(例如，见图 10.56 和图 10.58)。请预设一个参量，使得应用这个算法无论如何总能得到闭合形式的边界。

★10.43 对于下图所示的一维灰度横截面，给出逐步构建水坝的实现过程。在每一步骤画出显示水位和已构建大坝的横截面。



814 ~ 815

10.44 如果我们在式(10.6-4)中针对 T(而非针对 $-T$)进行测试，那么图 10.59(c)中的负 ADI 图像会是什么样子？

10.45 判断下列说法是否正确。对你的答案进行解释。

★(a) 只要物体正在运动，那么绝对 ADI 中的非零项的维数就会持续增加。

(b) 不管物体是否运动，正 ADI 中的非零项总是占据相同的面积。

(c) 只要物体正在运动，负 ADI 的非零项的维数就会持续增加。

10.46 假设在例 10.28 中将 x 轴方向的运动设为零。现在物体仅沿 y 轴以每帧 1 像素的速度运动 28 帧，然后瞬间转向，向相反方向运动 28 帧。在这些条件下，图 10.63 和图 10.64 会是什么样子？

★**10.47** 预置一个参量，证明当式(10.6-12)和式(10.6-13)中的 S_{1x} 和 S_{2x} 的符号相同时，速度分量 V_1 为正。

10.48 一家自动制药厂为质量控制目的，采用图像处理方法来测量药片的形状。该系统的分割步骤是以 Otsu 方法为基础的。由于检测线的速度很高，因此需要速率非常高的闪光来“停止”运动。照明灯较新时，照明灯投影均匀的光线。灯泡老化时，照明模式的退化根据下面的公式是时间和空间坐标的函数：

$$i(x, y) = A(t) - t^2 e^{-[(x-M)^2 + (y-N)^2]}$$

式中，(M, N)是观察区域的中心，t 是以月为单位增加的测试时间。灯泡是试验性的，且制造商并不了解灯光的 $A(t)$特性。所知道的全部是在灯泡的寿命期，$A(t)$总是大于上式中的负值，因为光照不能为负。观察表明，当灯泡较新时，Otsu 算法工作得很好，其光照模式在整个图像上近似于常数。然而，分割性能却随时间恶化。由于试验性灯泡格外昂贵，因此请你作为顾问在计算上帮助解决该问题，从而延长灯泡的寿命。给你很大的灵活性，可以在成像摄像机取景区域的边缘附近安装任何特殊标记或其他可视提示。请提出一个能让工程计划经理理解的详细解决方案。(提示：回顾 2.3.4 节中讨论的图像模型，并考虑使用一个其反射率已知的小目标。)

10.49 飞行中的子弹的速度可以用成像技术来估计。选择一种包括使用一台 TV 摄像机和可以曝光场景 k 秒的闪光灯的方法。子弹长 2.5 cm，宽 1 cm，速度范围为 700 ± 200 m/s。光学摄像机拍摄了一幅水平分辨率为 256×256 的数字图像，其中子弹占据了 10%。

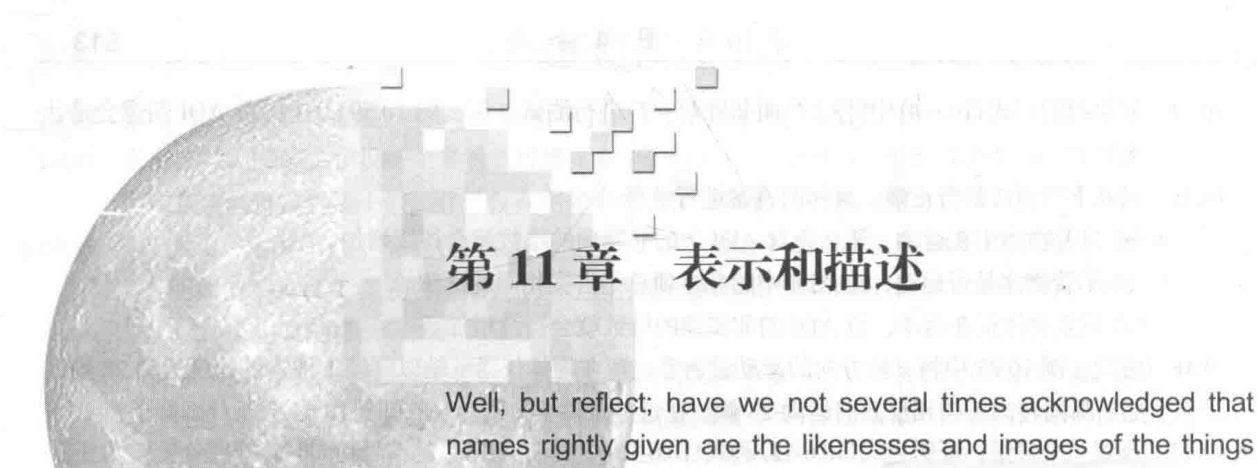
★(a) 为保证运动对图像造成的模糊不超过 1 像素，确定最大的 K 值。

(b) 确定每秒必须得到的帧数的最小值，以便保证在子弹穿过摄像机的视场时，至少得到两幅子弹的完整图像。

(c) 提出一种能自动地从帧序列中提取子弹的分割过程。

(d) 提出一种能自动地确定子弹速度的方法。

815
~
816



第 11 章 表示和描述

Well, but reflect; have we not several times acknowledged that names rightly given are the likenesses and images of the things which they name?

Socrates

引言

使用第 10 章中讨论的那些方法将一幅图像分割成多个区域后，分割后的像素集经常以一种适合于计算机进一步处理的形式来表示和描述。基本上，表示一个区域涉及两种选择：(1)我们可以根据其外部特征(其边界)来表示区域，或(2)根据其内部特征(如组成该区域的像素)表示区域。然而，选择表示方案仅是使得数据适用于计算机的一部分。下一项任务是基于所选择的表示来描述区域。例如，区域可由其边界表示，边界可用特征对其进行描述，如长度、连接端点的直线的方向，以及边界上凹陷的数量。

当我们关注的重点是形状特征时，可选择一种外部表示；而当关注的重点是内部属性如颜色和纹理时，可以选择一种内部表示。有时，需要同时使用这两种表示。无论哪种情形，选择用来作为描绘子的特征都应尽可能地对大小、平移和旋转不敏感。本章中讨论的多数描绘子满足一种或多种这样的特性。

11.1 表示

第 10 章中讨论的分割技术以沿着边界或包含在区域中的像素的形式产生原始数据。标准做法是使用某种方案将分割后的数据精简为便于描绘子计算的表示。本节讨论各种表示方法。

11.1.1 边界追踪

本章中讨论的几个算法要求一个区域的边界上的点以顺时针(或逆时针)方向排序。因此，我们通过介绍边界跟踪算法开始我们的讨论，该算法的输出是排过序的点的序列。我们假设：(1)处理的是二值图像，其目标和背景点分别标为 1 和 0；(2)图像已使用值为 0 的边界填充，因而消除了目标与图像边界合并的可能性。为方便起见，我们仅限于讨论单个区域。通过单独地处理各个区域，该方法可扩展到多个不相交的区域。

> 在继续阅读前，请读者复习一下 2.5.2 节和 9.5.3 节中的内容。

817～818

给定一个二值区域 R 或其边界，追踪 R 的边界或给定边界的算法由如下步骤组成：

1. 令起始点 b_0 为图像中左上角标记为 1 的点①。使用 c_0 表示 b_0 西侧的邻点［见图11.1(b)］。很明显，c_0 总是背景点。从 c_0 开始按顺时针方向考察 b_0 的 8 个邻点。令 b_1 表示所遇到的值为 1 的第一个邻点，并直接令 c_1(背景)是序列中 b_1 之前的点。存储 b_0 和 b_1 的位置，以便在步骤 5 中使用。
2. 令 $b = b_1$ 和 $c = c_1$［见图11.1(c)］。
3. 从 c 开始按顺时针方向行进，令 b 的 8 个邻点为 $n_1, n_2, \cdots, n_8$。找到标为 1 的第一个 n_k。
4. 令 $b = n_k$ 和 $c = n_{k-1}$。
5. 重复步骤 3 和步骤 4，直到 $b = b_0$ 且找到的下一个边界点为 b_1。

算法停止时，所找到的 b 点的序列就构成了排列后的边界点的集合。

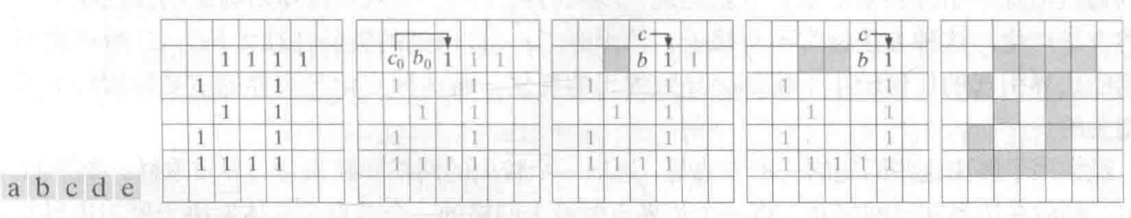


图11.1　边界追踪算法前几步的说明。下一个被处理的点标为黑色，而已被处理的点标为灰色，算法找到的点标为灰色方块

注意，步骤 4 中的 c 总是背景点，因为 n_k 是顺时针扫描时找到的第一个 1 值点。在 Moore[1968] 之后，有时该算法称为 Moore *边界追踪算法*。该算法的步骤 5 中规定的停止规则并不正确，因为第一次停止时，会再次遇到 b_0。正像我们马上要了解的那样，这会导致错误的结果。

图11.1 显示了刚才讨论的边界追踪算法的前几步。很容易检验，继续该过程将得到如图11.1(e)所示的正确边界，该边界中的点是一个顺时针方向排列的序列。

为了解该算法步骤 5 中规定的停止规则的必要性，考虑图11.2 中的边界。例如，不完全的毛刺去除(见 9.5.8 节有关毛刺的说明)会导致边界上侧的线段。从左上角的点开始，得到所示的步骤。在图 11.2(c)中，我们看到该算法已回到起始点。如果步骤因为再次到达起始点而停止，那么显然不会找到剩余的边界。使用步骤 5 中的停止规则可允许算法继续，并且显示图 11.2 中找到的整个边界很简单。

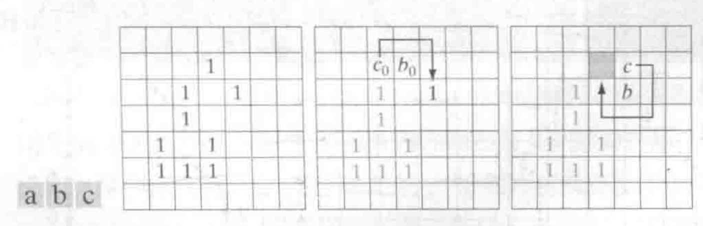


图 11.2　当再次遇到起始点 b_0 时，边界追踪算法满足停止规则导致的错误结果的说明

如果给定一个区域而非其边界(就像在前述说明中那样)，那么边界追踪算法会工作得很好。也就是说，该过程提取一个二值区域的外边界。如果目的是找到一个区域中的孔洞的边界(这种边界称为该区域的*内边界*)，一种简单的方法是提取这些孔洞(见 9.5.9 节)，并将它们当作 0 值背景上的 1 值区域来处理。对这些区域应用边界追踪算法将得到原始区域的内边界。

① 在本章后面我们会看到，边界中的左上角点有一个重要性质，即该边界的一个多边形近似在该位置有一个凸顶点。此外，该点的左上邻点可保证为背景点。这些属性使得该点成为开始边界追踪算法的一个较好的"标准"点。

818 ~ 819

我们已经简单地说明了基于顺时针方向跟踪一个边界的算法。事实上，我们会遇到按如下假设来进行表述的算法：边界点已在某个方向上排序。在以下几节中，我们将交替使用两个方向来帮助读者熟悉这两种算法。

11.1.2 链码

链码用于表示由顺次连接的具有指定长度和方向的直线段组成的边界。通常，这种表示基于这些线段的 4 连接或 8 连接。每个线段的方向使用一种数字编号方案编码，如图11.3所示。以这种方向性数字序列表示的编码称为佛雷曼链码。

数字图像通常以网格形式来获取并处理，在这种网格形式中，x 和 y 方向的间距相等，所以链码可以通过追踪一个边界来产生，也就是说，以顺时针方向，并对连接每对像素的线段赋予一个方向的方法产生。这种方法通常不可接受，原因有二：(1)得到的链码往往太长，(2)噪声或不完美分割沿边界引起的任何较小干扰都会导致编码的变化，而这种变化与边界的主要形状特征可能是不相关的。

经常用于解决这些问题的一种方法是，选取一个较大的网格间距来对边界重取样，如图 11.4(a)所示。然后在边界穿过网格时，将一个边界点赋给大网格的一个节点，具体取决于原始边界与该节点的接近程度，如图 11.4(b)所示。再后，按这种方法得到的重取样边界可由一个 4 链码或 8 链码表示。图 11.4(c)显示了由 8 方向链码表示的粗略边界点。从 8 链码转换为 4 链码或从 4 链码转换为 8 链码很简单(见习题 2.12 和习题 2.13)。图 11.4(c) 中的起始点位于边界的左上角处，它给出了链码 0766…12。如所期望的那样，结果编码表示的精度取决于取样网格的间距。

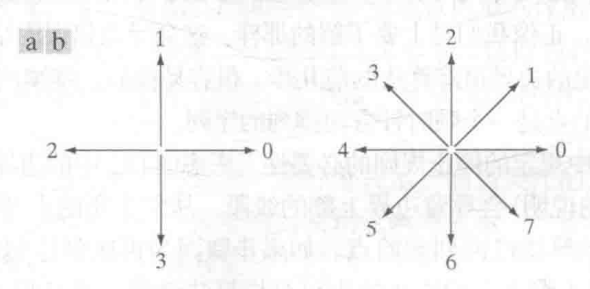


图 11.3 (a) 4 方向链码和(b) 8 方向链码的方向编号

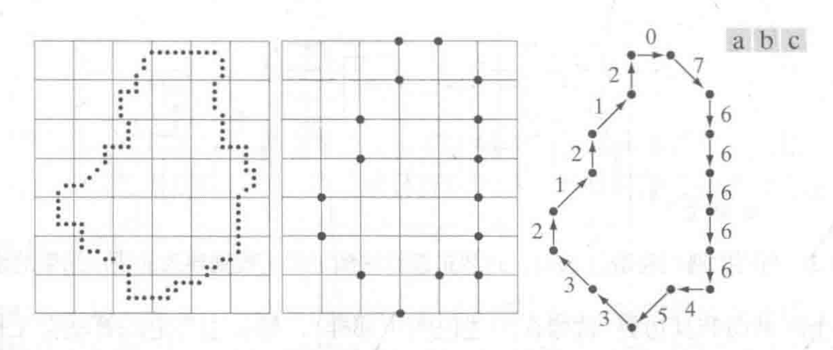


图 11.4 (a)叠加有重取样网格的数字边界；(b)重取样的结果；(c) 8 方向链码边界

边界的链码取决于起始点。然而，链码可以通过一个简单的过程关于起始点归一化，过程如下：我们简单地将链码视为方向号码的一个循环序列，并重新定义起始点，以便得到号码序列的最小整数值。我们也可以针对旋转归一化(使用图 11.3 中方向的整数倍的角度)，方法是使用链码的一次差分来替代链码本身。这个差分是通过计算链码中分隔两个相邻像素的方向变化的数(在图11.3 中按逆时针方

819 ~ 821

向)得到的。例如，4 方向链码 10103322 的一次差分为 3133030。如果我们把链码作为循环序列来关于起始点归一化，则差分的第一个元素是通过使用链码的最后一个元素和第一个元素间的转变来计算得到的。这里，结果是 33133030。改变重取样网格的大小，可实现尺寸归一化。

只有对于旋转和尺度变化而边界本身不变(再次强调，按图11.3 中方向的整数倍角度)时，这些归一化才是准确的，实际中很少出现这种情形。例如，同一物体按两个不同方向数字化后，通常会有不同的边界形状，不相似程度与图像的分辨率成正比。按其长度与数字化图像中像素间距离成比例来选择链码，和/或沿将被编码物体的主轴(见11.2.2 节中的讨论)或本征轴(见11.4 节中的讨论)来选定重取样网格的方向，可降低这种影响。

例 11.1　佛雷曼链码及它的某些变化。

图 11.5(a)显示了小镜片中嵌入环形笔画后的图像，该图像是大小为 570×570 的 8 比特灰度图像。本例的目的是得到图 11.5(a)的最小值整数的佛雷曼链码和最大物体外部边界的一次差分。因为感兴趣物体镶嵌在小镜片中，提取其边界的结果将是一条不能描述该物体的普通形状的带噪曲线。图 11.5(b)显示了使用一个大小为 9×9 的均值模板平滑原图像后的结果，图 11.5(c)是采用 Otsu 算法得到的全局阈值对该图像进行阈值处理后的结果。注意，区域数量已减少到两个(其中一个区域是一个点)，从而大大简化了该问题。

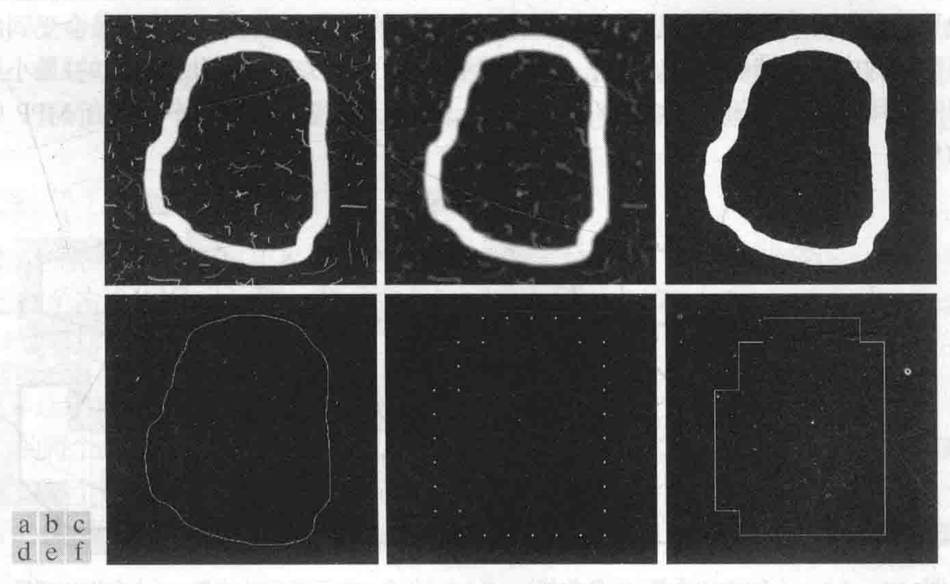


图 11.5　(a)带噪声图像；(b)使用一个大小为 9×9 的均值模板平滑图像后的结果；(c)使用 Otsu 算法对平滑图像进行阈值处理后的结果；(d)图(c)的最长外部边界；(e)子采样边界(为清楚起见，点已被放大)；(f)使用直线连接图(e)中的点后的结果

图11.5(d)是图11.5(c)中最大区域的外边界。直接获得该边界的链码会生成一个变化较小的较长序列，且不能表示该边界的形状。如本节之前提到的那样，为降低可变性，在得到其边界的链码前，通常需要对该边界重取样。图11.5(e)是以节点间距为 50 像素(约为图像宽度的 10%)的网格对边界重取样后的结果，图11.5(f)是使用直线连接所得顶点后的结果。这一简单近似保留了原始边界的主要特征。

简化后的边界的 8 方向佛雷曼链码是

0 0 0 0 6 0 6 6 6 6 6 6 6 6 4 4 4 4 4 4 2 4 2 2 2 2 2 0 2 2 0 2

该边界的起始点位于子取样网格中的坐标(2, 5)处。这是图11.5(f)中的左上角点。这种情形下，该链码的最小值整数刚好与如下链码的最小值整数相同：

821~822

00006066666666444444242222202202

两个链码的一次差分均为

00062600000000600000626000006206 26

使用这两种链码之一表示边界时，可有效地减少用于存储该边界的数据量。此外，如我们在 11.2 节中讨论的那样，使用码数可提供一种统一的方法来分析边界的形状。最后，记住子采样边界可以由上述链码之一来恢复。

11.1.3 使用最小周长多边形的多边形近似

数字边界可以用多边形以任意精度来近似。对于一条闭合边界，当多边形的边数等于边界上的点数时，这种近似会变得很精确，此时，每对相邻的点定义了多边形的一条边。多边形近似的目的是使用数量尽可能少的线段来获取给定边界的基本形状。通常，该问题并不容易求解，且求解过程会转化为耗时的迭代搜索。然而，适度复杂的多边形近似技术还是很适合于图像处理任务。在这些近似技术中，最有效的一种技术是使用最小周长多边形(MPP)来表示边界，详见如下讨论中的定义。

基础

产生计算 MPP 的算法的一种直观方法是，把图11.6(b)那样的一组连接单元的集合封入图11.6(a)中的一条边界。我们可以将该边界想象为一个橡皮条。允许橡皮条收缩时，橡皮条会受到由这些单元定义的边界区域的内、外墙的约束。最终，该收缩会产生一个(关于这种几何排列的)最小周长的多边形，它被限制在用单元条封闭的区域内，如图11.6(c)所示。注意，在该图中，所有 MPP 的顶点与内墙或外墙的角点一致。

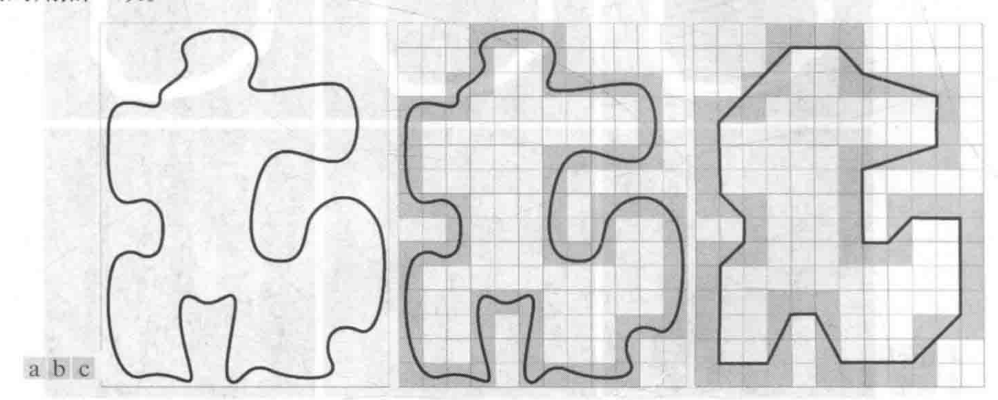


图11.6 (a)一个物体的边界(黑色曲线)；(b)由(灰色)单元围成的边界；(c)允许边界收缩得到的最小周长多边形。多边形的顶点是灰色区域的内墙角点和外墙角点

单元的大小决定着多边形近似的精度。在极限情况下，如果每个(方形)单元的大小对应于边界上的一个像素，则边界和 MPP 近似值之间的每个单元中的误差至多为 $\sqrt{2}d$，其中 d 是像素间的最小可能距离(即由原始取样边界的分辨率决定的像素间的距离)。通过强迫多边形近似中的每个单元的中心为原始边界上的对应像素，可以将该误差减半。在给定的应用中，我们的目的是使用合适的最大可能单元大小，以最少的顶点数来产生 MPP。在这一节中，我们的目的是阐明寻找这些 MPP 顶点的过程。

刚才讨论的单元方法将由原始边界包围的物体的形状简化为图 11.6(b)中由灰墙围成的区域。图 11.7(a)将这一形状显示为暗灰色。我们看到，其边界由 4 连接的直线段组成。假设我们按顺时针方向追踪这条边界，追踪时遇到的每个转向要么是一个凸顶点，要么是一个凹顶点，其中顶点的角是 4 连接边界的内角。凸顶点和凹顶点在图 11.7(b)中分别显示为白点和黑点。注意，这些顶点是图 11.7(b)

822 ~ 824

中浅灰色边界区域的内墙的顶点，暗灰色区域中的每个凹顶点(黑点)在浅灰色墙中有一个对应的“镜像”顶点，镜像顶点位于这个凹顶点的对角处。图 11.7(c)显示了所有凹顶点的镜像顶点，为方便参考，图中叠加了图11.6(c)中的 MPP。我们看到，MPP 的顶点要么与内墙中的凸顶点(白点)一致，要么与外墙中的凹顶点的镜像顶点(黑点)一致。稍加思考就会发现，只有内墙的凸顶点和外墙的凹顶点才能成为 MPP 的顶点。因此，我们的算法就只需要关注这些顶点。

凸顶点是一组三个点的中心点，它定义了范围 $0° < \theta < 180°$ 内的一个角。类似地，凹顶点定义了范围 $180° < \theta < 360°$ 内的一个角。180° 角定义了一个退化顶点(一条直线)，退化顶点不可能是一个 MMP 顶点。0° 或 360° 角涉及折回一条路径，在本讨论中这是一个无效的条件。

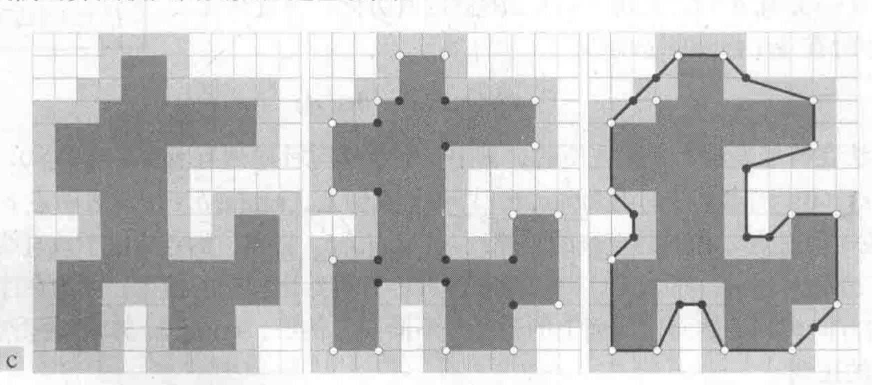


图11.7　(a)由单元围成的原始边界(见图11.6)得到的区域(暗灰色)；(b)按逆时针方向追踪暗灰色区域的边界得到的凸顶点(白点)和凹顶点(黑点)；(c)移到边界区域的外墙中对角镜像位置的凸顶点(黑点)，凹顶点保持不变。为便于参考，图像中叠加了 MPP(黑色边界)

MPP 算法

前一段描述的围成一条数字边界的单元集合称为单元组合体。我们假设所考虑的边界本身不相交，这将产生简单连接的单元组合体。基于这些假设，并令 W(白)和 B(黑)分别表示凸顶点和镜像凹顶点，我们将观察结果说明如下：

1. 由简单连接的单元组合体为界的 MPP 是非自相交的。
2. MPP 的每个凸顶点都是一个 W 顶点，但并非边界的每个 W 顶点都是 MPP 的一个顶点。
3. MPP 的每个镜像凹顶点都是一个 B 顶点，但并非边界的每个 B 顶点都是 MPP 的一个顶点。
4. 所有的 B 顶点要么在 MPP 上，要么在 MPP 外；所有的 W 顶点要么在 MPP 上，要么在 MPP 内。
5. 单元组合体中包含的顶点序列的最左上角顶点，总是 MPP 的一个 W 顶点。

这些断言已被正式证明(Sklansky et al.[1972], Sloboda et al.[1998]和 Klette and Rosenfeld[2004])。然而，它们的正确性对于我们的目的来说非常明显(见图 11.7)，所以此处对它们的证明不再赘述。与图11.7中暗灰色区域的顶点的角度不同的是，MPP 顶点维持不变的角度不必是 90°的倍数。

在下面的讨论中，我们需要计算点的三元组的方向。考虑点的三元组(a, b, c)，并令这些点的坐标为 $a=(x_1,y_1)$，$b=(x_2,y_2)$ 和 $c=(x_3,y_3)$。如果我们以矩阵的行来安排这些点，

$$\boldsymbol{A}=\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \tag{11.1-1}$$

则由基本矩阵分析有

$$\det(A) = \begin{cases} > 0, & \text{如果}(a,b,c)\text{是一个逆时针序列} \\ = 0, & \text{如果点是共线的} \\ > 0, & \text{如果}(a,b,c)\text{是一个顺时针序列} \end{cases} \tag{11.1-2}$$

其中 det(A)是 A 的行列式。根据这个公式，逆时针或顺时针方向运动与右手坐标系有关(见 2.4.2 节的脚注)。例如，使用该图像坐标系(见图 2.18)，原点位于左上角，x 轴的正方向垂直向下，y 轴的正方向水平向右，序列 $a = (3, 4)$, $b = (2, 3)$和 $c = (3, 2)$是逆时针方向，并且在代入式(11.1-2)中时有 $\det(A) > 0$。故定义

> 假设使用图 2.18(b)中定义的坐标系，当按逆时针方向追踪一个多边形的边界时，追踪方向右侧的所有点位于多边形的外部，追踪方向左侧的所有点位于多边形的内部。

$$\mathrm{sgn}(a,b,c) \equiv \det(A) \tag{11.1-3}$$

来描述该算法在符号表示上是很方便的，其中，对于逆时针序列有 $\mathrm{sgn}(a, b, c) > 0$，对于顺时针序列有 $\mathrm{sgn}(a, b, c) < 0$，当点共线时有 $\mathrm{sgn}(a, b, c) = 0$。几何上，$\mathrm{sgn}(a, b, c) > 0$ 表明点 c 位于线对(a, b)的正侧(即 c 位于通过点 a 和 b 的直线的正侧)。若 $\mathrm{sgn}(a, b, c) < 0$，则点 c 位于该直线的负侧。如果使用序列(c, a, b)或(b, c, a)，那么式(11.1-2)和式(11.1-3)可给出相同的结果，因为在序列中追踪的方向与(a, b, c)的方向相同。然而，几何上的解释是不同的。例如，$\mathrm{sgn}(c, a, b) > 0$ 表明点 b 位于通过点 c 和 a 的直线的正侧。

为了给 MPP 算法准备数据，我们形成一个列表，该列表中的行是每个顶点的坐标和一个表示顶点是 W 顶点还是 B 顶点的附加元素。如图11.7(c)所示，凹顶点被镜像为按顺序排列的顶点很重要[①]，第一个顶点是左上角的顶点，由性质 5 可知它是 MPP 的 W 顶点。令 V_0 表示这个顶点。我们假设顶点按逆时针方向排列。寻找 MPP 的算法使用两个“爬行”点：一个白色的爬行点(W_C)和一个黑色的爬行点(B_C)。W_C 沿凸顶点(W)爬行，B_C 沿镜像凹顶点(B)爬行。这两个爬行点、找到的最后一个 MPP 顶点和正被考察的顶点，都是实现该过程所必需的顶点。

算法首先令 $W_C = B_C = V_0$ (回忆可知，V_0 是一个 MPP 顶点)。然后，在算法的任何步骤中，令 V_L 表示找到的最后一个 MPP 顶点，令 V_k 表示正在考察的当前顶点。在 V_L, V_k 和两个爬行点之间存在如下三个条件之一：

(a) V_k 位于通过(V_L, W_C)的直线的正侧，即 $\mathrm{sgn}(V_L, W_C, V_k) > 0$。

(b) V_k 位于通过(V_L, W_C)的直线的负侧，或 V_k 与(V_L, W_C)共线，即 $\mathrm{sgn}(V_L, W_C, V_k) \le 0$。同时，$V_k$ 位于通过(V_L, B_C)的直线的正侧，或 V_k 与(V_L, B_C)共线，即 $\mathrm{sgn}(V_L, B_C, V_k) \ge 0$。

(c) V_k 位于通过(V_L, B_c)的直线的负侧，即 $\mathrm{sgn}(V_L, B_C, V_k) < 0$。

如果条件(a)成立，则下一个 MPP 顶点是 W_C，并且我们令 $V_L = W_C$；然后，我们令 $W_C = B_C = V_L$ 来重新初始化该算法，并在 V_L 之后的下一个顶点继续执行算法。

如果条件(b)成立，则 V_k 变成一个候选的 MPP 顶点。在这种情形下，如果 V_k 是凸顶点(即一个 W 顶点)，则令 $W_C = V_k$；否则，我们置 $B_C = V_k$。然后，我们使用列表中的下一个顶点继续执行算法。

如果条件(c)成立，则下一个 MPP 顶点是 B_C，并且令 $V_L = B_C$；然后，我们令 $W_C = B_C = V_L$ 来重新初始化该算法，并用 V_L 之后的下一个顶点继续执行算法。

再次到达第一个顶点时，算法结束，此时算法已处理了多边形中的所有顶点。算法找到的 V_L 顶点是 MPP 的顶点。业已证明，该算法可找到一个被简单连接的单元组合体围成的多边形的所有 MPP 顶点(Sloboda et al.[1998], Klette and Rosenfeld[2004])。

826 ~ 827

① 一个边界的顶点可按算法追踪边界的方向排序，例如，使用 11.1.1 节中描述的算法来追踪边界。

例 11.2　MPP 算法示例。

一个手算例子有助于阐明前面的概念。考虑图11.7(c)中的顶点。在我们的图像坐标系中，网格左上角的点的坐标是(0, 0)。假设网格间距为 1，则(逆时针)顶点列表中的前几行为

V_0	(1, 4)	W
V_1	(2, 3)	B
V_2	(3, 3)	W
V_3	(3, 2)	B
V_4	(4, 1)	W
V_5	(7, 1)	W
V_6	(8, 2)	B
V_7	(9, 2)	B

列表的第一个元素总是我们的第一个 MPP，因此首先令 $W_C = B_C = V_0 = V_L = (1,4)$ 。下一个顶点是 $V_1 = (2, 3)$ 。由符号函数计算得 $\text{sgn}(V_L, W_C, V_1) = 0$ 和 $\text{sgn}(V_L, B_C, V_1) = 0$ ，因此，条件(b)成立。因为 V_1 是一个 B(凹)顶点，我们令 $B_C = V_1 = (2,3)$ 。W_C 保持不变。在这一阶段，爬行点 W_C 位于(1, 4)处，爬行点 B_C 位于(2, 3)处，而 V_L 仍位于(1, 4)处，因为未找到新的 MPP 顶点。

下面，我们考虑 $V_2 = (3,3)$ 的情形。符号函数的值是 $\text{sgn}(V_L,W_C,V_2) = 0$ 和 $\text{sgn}(V_L,B_C,V_2) = 1$ ，所以算法的条件(b)再次成立。因为 V_2 是一个 W(凸)顶点，我们令 $W_C = V_2 = (3,3)$ 。在这一阶段，爬行点位于 $W_C = (3,3)$ 和 $B_C = (2,3)$ 处；V_L 保持不变。

下一个顶点是 $V_3 = (3, 2)$ 。符号函数的值是 $\text{sgn}(V_L, W_C, V_3) = -2$ 和 $\text{sgn}(V_L, B_C, V_3) = 0$ ，因此条件(b)再次成立。因为 V_3 是一个 B 顶点，我们将黑色爬行点更新为 $B_C = (3,2)$ 。爬行点 W_C 保持不变，V_L 保持不变。

下一个顶点是 $V_4 = (4,1)$ ，且有 $\text{sgn}(V_L,W_C,V_4) = -3$ 和 $\text{sgn}(V_L,B_C,V_4) = 0$ ，因此条件(b)再次成立。因为 V_4 是一个白色顶点，我们将白色爬行点更新为 $W_C = (4,1)$ 。黑色爬行点 B_C 保持在(3, 2)处，V_L 仍位于(1, 4)处。

下一个顶点是 $V_5 = (7,1)$ ，且 $\text{sgn}(V_L,W_C,V_5) = 9$ ，因此条件(a)成立，我们令 $V_L = W_C = (4,1)$ 。因为找到了一个新的 MPP 顶点，因此我们令 $W_C = B_C = V_L$ 来重新初始化算法，并使用新近找到的 V_L 后的下一个顶点开始算法。下一个顶点是 V_5，所以我们再次访问它。

使用 $V_5 = (7,1)$ 和 V_L , W_C , B_C 的新值，我们得到 $\text{sgn}(V_L, W_C, V_5) = 0$ 和 $\text{sgn}(V_L, B_C, V_5) = 0$ ，故条件(b)成立。因此，我们令 $W_C = V_5 = (7,1)$ ，因为 V_5 是一个 W 顶点。

下一个顶点是 $V_6 = (8,2)$ ，且 $\text{sgn}(V_L, W_C, V_6) = 3$ ，所以条件(a)成立。这样，我们令 $V_L = W_C = (7,1)$ ，并通过令 $W_C = B_C = V_L$ 来重新初始化算法。

因为在(7, 1)处重新初始化，故下一个要考虑的顶点再次为 $V_6 = (8,2)$ 。按上面的方法继续处理该顶点及剩下的顶点，可得到图11.7(c)中的 MPP 顶点。如早些时候提到的那样，位于(2, 3)处和(3, 2)处及位于(13, 10)处右下方的镜像 B 顶点是共线的，尽管它们位于 MPP 的边界上，但不能视为 MPP 的顶点。这是合适的，算法的确没有这样检测它们。

例 11.3　MPP 算法的应用。

图 11.8(a)是一幅大小为 566×566 的枫叶二值图像，图 11.8(b)是其 8 连接边界。图11.8(c)到(e)中的序列图像显示了该边界使用大小分别为 2, 3, 4, 6, 8, 16 和 32 的方形单元组合体的 MPP 表示(每幅图中的顶点都用直线连接起来形成一个闭合的边界)。枫叶有两个主要特征：茎和三个主要的圆裂片。如图 11.8(f)所示，当单元尺寸大于 4×4 时，茎开始消失。而即使是对于大小为 16×16 的单元尺寸，仍完好地保留了

827 ~ 828

三个主要的圆裂片，如图 11.8(h)所示。然而，在图 11.8(i)中，我们看到当单元尺寸增大到 32×32 时，这一明显的特征也几乎消失。

原始边界[见图 11.8(b)]中的点数是 1900。图 11.8(c)到(i)中的顶点数分别是 206, 160, 127, 92, 66, 32 和 13。尽管将数据量降低了 90%以上，但有 127 个顶点的图 11.8(e)仍保留了原始边界的所有主要特征，。由此我们可以看出使用 MPP 来表示边界的明显优势。MPP 的另一个优点是可以进行边界平滑。如前节中解释的那样，当我们使用链码来表示一条边界时，通常需要进行边界平滑。

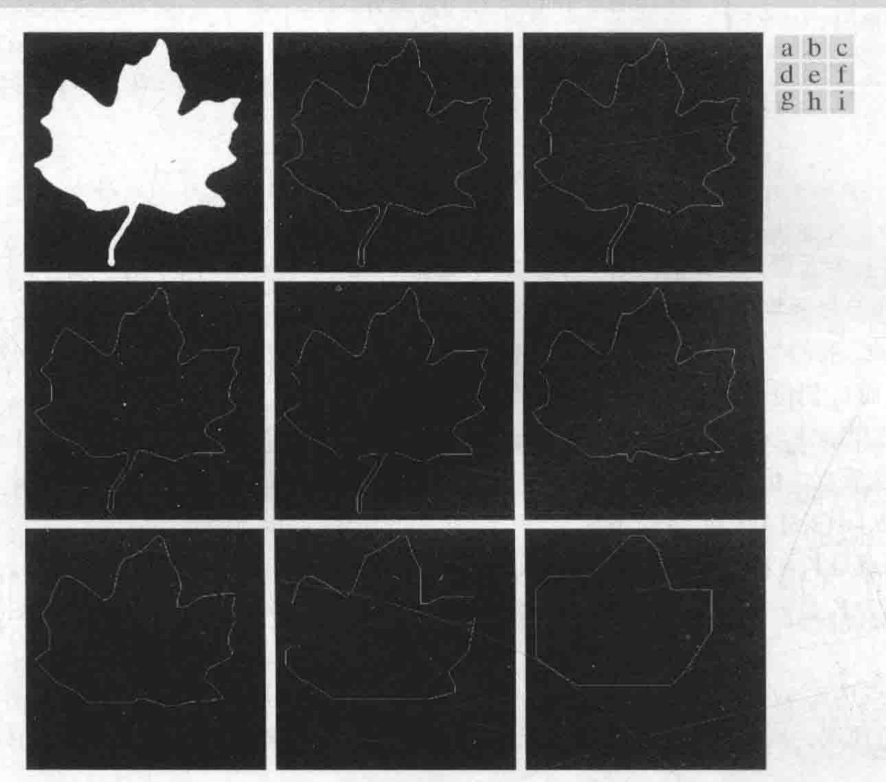


图 11.8 (a)大小为 566×566 的二值图像；(b)8 连接边界；(c)~(i)使用大小分别为 2, 3, 4, 6, 8, 16 和 32 的方形单元得到的 MPP(为显示方便，用直线将这些顶点连接起来)。图(b)中边界的顶点数量为 1900。图(c)到图(i)中的顶点数量分别为 206, 160, 127, 92, 66, 32 和 13

11.1.4 其他多边形近似方法

有时，可以用概念上比前节中讨论的 MPP 算法更简单的方法来进行多边形近似。本节讨论两种这样的方法。

聚合技术

基于平均误差或其他准则的聚合技术已应用于多边形近似问题。一种方法是沿一条边界来聚合一些点，直到拟合这些聚合点的直线的最小均方误差超过某个预设的阈值。当这种条件出现时，存储该直线的参数，将误差设为零，并且重复该过程，沿边界聚合新的点，直到该误差再次超过预设的阈值。这一过程结束后，相邻线段的交点就构成多边形的顶点。这种方法的主要难点是，得到的近似顶点并不总是对应于原始边界中的形变(如拐角)，因为在误差超过阈值前，不会开始画一条新的直线。例如，如果沿着一条长的直线追踪，且遇到了一个拐角，那么通过该拐角的许多点(取决于阈值)在误差超过阈值前将被丢弃。然而，(下面讨论的) 随同聚合技术一起的分裂技术可用于缓解这一困难。

828~830

分裂技术

分裂边界线段的一种方法是将线段不断地细分为两部分，直到满足规定的准则为止。例如，一个要求可能是：一条边界线段到连接其两个端点的直线间的最大垂直距离不超过一个预设的阈值。如果准则满足，则与直线有着最大距离的点就成为一个顶点，这样就将初始线段分成了两条子线段。这种方法在寻找变化显著的点时具有优势。对一条闭合边界，最好的起始点通常是边界上的两个最远点。例如，图11.9(a)显示了一个物体的边界，图11.9(b)显示了这条边界关于其最远点的细分。标记为 c 的点是顶部边界线段到直线 ab 的(垂直距离)最远点。类似地，点 d 是底部线段上的最远点。图11.9(c)显示了使用分裂过程后的结果，所用阈值为直线 ab 的长度的 0.25 倍。由于新边界线段上没有(到其对应的直线段的)垂直距离超过该阈值的点，故该过程终止于图11.9(d)中的多边形。

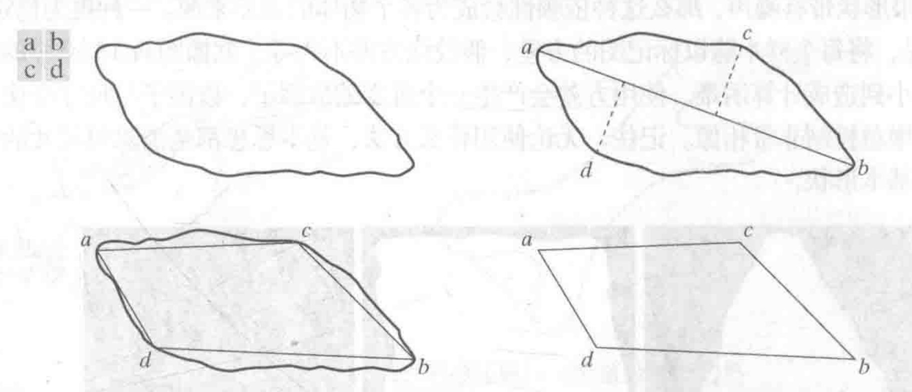


图 11.9　(a)原始边界；(b)边界被分割为基于端点的线段；(c)顶点连接；(d)得到的多边形

11.1.5　标记图

标记图是边界的一维函数表示，它可以使用各种方式来生成。一种最简单的方法是以角度函数的形式画出质心到边界的距离，如图11.10所示。然而，不管如何生成标记图，基本概念都是将边界表示简化为描述起来可能比原始二维边界更简单的一维函数。

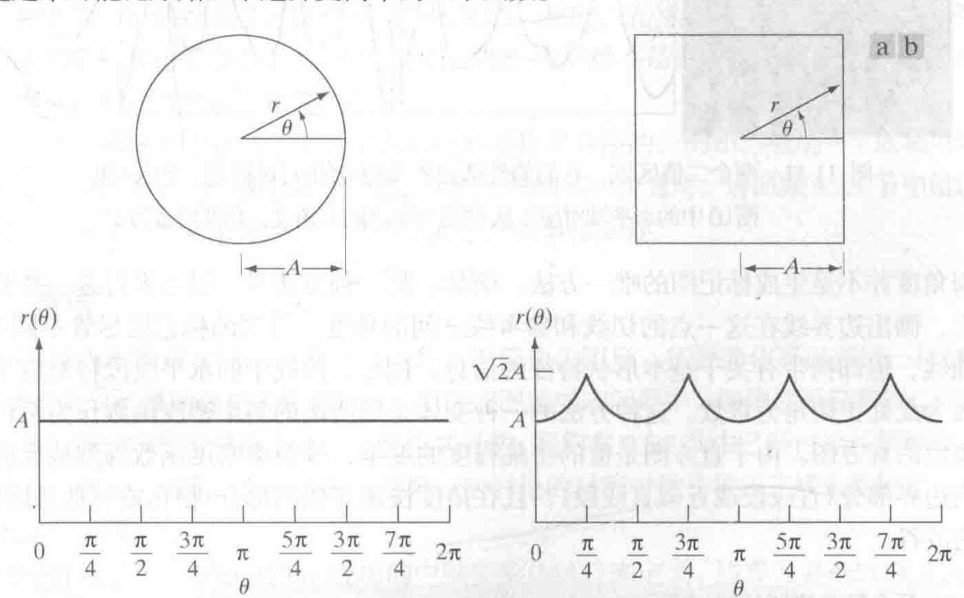


图 11.10　距离与角度标记图。在图(a)中，$r(\theta)$为常量。在图(b)中，标记图由重复出现的模式 $r(\theta) = A\sec\theta, 0 \leq \theta \leq \pi/4$ 和 $r(\theta) = A\csc\theta, \pi/4 < \theta \leq \pi/2$ 构成

830～831

由刚才描述的方法所生成的标记图是平移不变的，但它们的确取决于旋转和缩放。通过寻找一种选取相同起始点来生成标记图的方法，可实现关于旋转的归一化，而不用考虑形状的方向。这样做的一种方法是，将距质心最远的点选为起始点，假设该点对于我们感兴趣的每个形状而言都是唯一的。另一种方法是在本征轴上选取（见 11.4 节）距质心最远的点。这种方法的计算量更大，但更为稳定，因为本征轴的方向是使用所有轮廓点确定的。当然，还有一种方法是获得该边界的链码，然后使用 11.1.2 节中讨论的方法，假定得到的编码足够粗糙，以至于旋转不会影响到它的曲率。

根据关于两个轴缩放一致且以等间隔θ取样的假设，形态大小的变化会导致相应标记图的幅值变化。对此进行归一化的一种方法是，对所有函数进行缩放，以便它们具有相同的值域，譬如[0, 1]。这种方法的主要优点是简单，但也有潜在的严重缺陷，即对整个函数的缩放只依赖于两个值：最小值和最大值。如果形状带有噪声，那么这种依赖性会成为各个物体的误差来源。一种更为稳定(但计算量更大)的方法是，将每个样本除以标记图的方差，假设该方差不为零［就像图11.10(a)中的情形那样］或该方差不会小到造成计算困难。使用方差会产生一个可变缩放因子，该因子与尺寸变化成反比，工作起来与自动增益控制非常相像。记住，无论使用什么方法，基本思想都是消除对尺寸的依赖性，同时保持波形的基本形状。

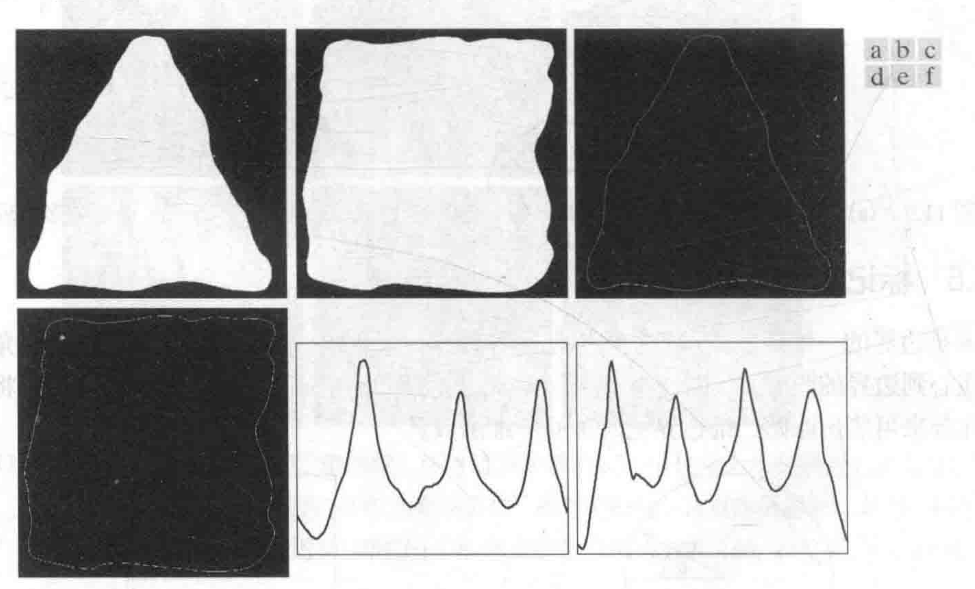


图 11.11　两个二值区域，它们的外部边界及相应的 $r(\theta)$ 标记。图(e)和图(f)中的水平轴对应于从 0°到 360°之间的角度，角度增量为 1°

距离与角度并不是生成标记图的唯一方法。例如，另一种方法是，沿边界行进，并对应于边界上的每个点，画出边界线在这一点的切线和参考线之间的角度。得到的标记图尽管不同于图 11.10 中的 $r(\theta)$曲线，但却携带有关于基本形状特征的信息。例如，曲线中的水平线段将对应于沿该边界的直线，因为此处正切角为常数。这种方法的一种变体是用所谓的斜率密度函数作为标记图。该函数是正切角值的直方图。由于直方图是值的密集程度的度量，故斜率密度函数强烈地反应了具有恒定正切角的边界部分(直线段或近似直线段)，且在角度快速变化的部分(拐角或其他急剧弯曲位置)存在较深的波谷。

例 11.4　两个简单物体的标记图。

831 ~ 832

图 11.11(a)和(b)显示了两个二值物体，图 11.11(c)和(d)是它们的边界。图 11.11(e)和(f)中相应的 $r(\theta)$ 标记图以 1°增量在 0°到 360°的范围内变化。标记图中突出峰值的数量足以区分两个物体的形状。

11.1.6　边界线段

将边界分解为线段通常是有很用的。分解降低了边界的复杂性，从而简化了描述过程。当边界线包含一个或多个携带形状信息的明显凹度时，这种方法尤其有吸引力。此时，使用由边界所围成区域的凸壳就成为边界鲁棒分解的有力工具。

如 9.5.4 节中定义的那样，一个任意集合 S 的凸壳 H 是包含 S 的最小凸集。差集 $H-S$ 称为集合 S 的凸缺 D。为了解如何使用这些概念来把边界分成有意义的线段，考虑图11.12(a)，图中显示了一个物体(集合 S)及其凸缺(阴影区)。区域边界可以按如下方式来分割：追踪 S 的轮廓，并标记进入或离开一个凸缺的转变点。图 11.12(b)显示了这种情况下的结果。注意，在原理上，这种方案与区域大小和方向无关。

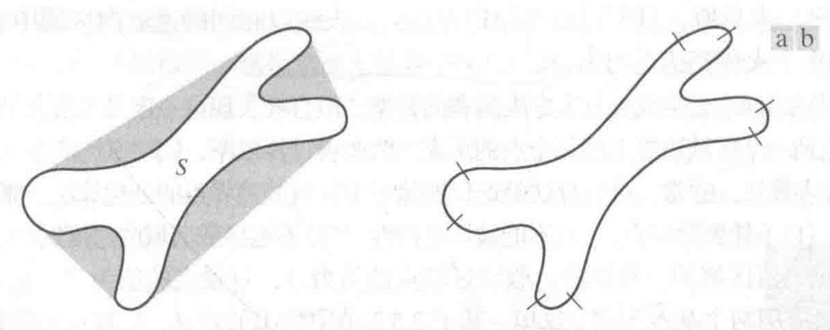


图 11.12　(a)区域 S 及其凸缺(阴影区域)；(b)被分割的边界

实际上，由于数字化、噪声和分割的变形的影响，数字边界往往是不规则的。这些影响通常会导致在整个边界上有着随机散布的无意义的小凸缺。与其试图通过后处理分出这些不规则边界，不如在对边界分割前用普通方法平滑边界。有很多方法可以做到这一点。一种方法是追踪边界，并使用一个像素沿该边界的 k 个相邻像素的平均坐标代替这个像素的坐标。这种方法适用于较小的不规则边界，但它耗费时间且难以控制。较大的 k 值会导致过度平滑，而较小的 k 值在某些边界线段可能会使平滑不够充分。一种更为稳定的技术是，在找到一个区域的凸缺前，先使用多边形近似。我们感兴趣的多数数字边界都是简单的多边形(回忆 11.1.3 节可知，这些多边形都不是自交叉的)。Graham and Yao[1983]给出了查找此类多边形的凸壳的一种算法。

凸壳及其凸缺的概念对于描述整个区域及其边界同样是有用的。例如，描述一个区域可以基于该区域的面积及其凸缺的面积、凸缺中的分量数、这些分量的相对位置等。请回顾 9.5.4 节中给出的一种寻找凸壳的形态学算法。本章末尾给出的参考文献包含有其他表述。

11.1.7　骨架

表示平面区域内结构形状的一种重要方法是将它简化为图形。这种简化可以通过一种细化(也称为骨架化)算法得到该区域的骨架来实现。在大范围的图像处理问题中，细化过程起着核心作用，从印制电路板的自动检测到空气过滤器中的石棉纤维的计数。我们在 9.5.7 节中已经讨论过使用形态学来骨架化的基础知识。然而，如那一节中解释的那样，所讨论的过程对保持骨架连接未做规定。这里给出的算法可纠正这一问题。

一个区域的骨架可以用 Blum[1967]提出的中轴变换(MAT)来定义。边界为 B 的区域 R 的 MAT 如下所示。对 R 中的每个点 p，我们在 B 中找到与其最接近的邻点。如果 p 有多个这样的邻点，则认为 p 属于 R 的中轴(骨架)。“最接近”(及得到的 MAT)这一概念取决于距离的定义(见 2.5.3 节)。图11.13显示了使用欧氏距离的一些例子。使用 9.5.7 节中的最大圆盘可得到相同的结果。

832～834

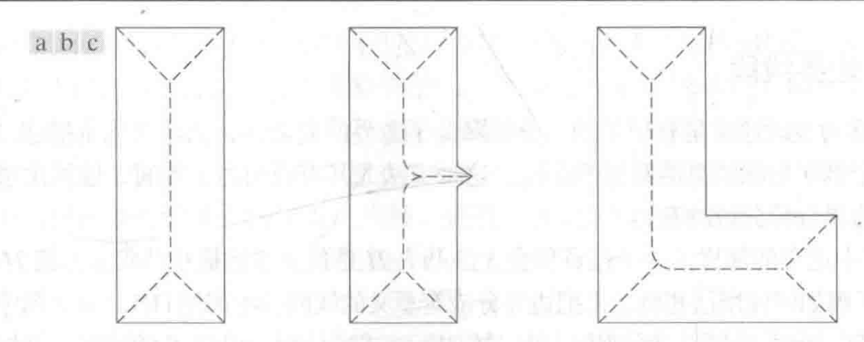


图 11.13　三个简单区域的中轴(虚线)

区域的 MAT 有一个直观的定义，该定义基于所谓的“草原之火概念”。我们将一个图像区域想象为由干草组成的平坦大草原，且假设沿草原边界点火。火线以相同的速度向区域中心推进。区域的 MAT 是同一时间多个火线到达点的集合。

尽管一个区域的 MAT 会生成直觉上令人满意的骨架，但直接实现这一定义需要大量的计算。实现涉及计算每个内部点到一个区域边界上的每个点的距离。为改善计算效率，同时试图产生一个区域的中轴表示，人们提出了许多算法。通常，这些算法是迭代删除一个区域的边界点的细化算法，删除这些点时要满足如下约束条件：(1)不能删除端点，(2)不能破坏连接性，(3)不能导致区域的过度腐蚀。

本节给出细化二值区域的一种算法。假设区域点的值为 1，背景点的值为 0。这种方法是对给定区域的边界点连续应用两个基本步骤，这里，基于 2.5.2 节中给出的定义，边界点是值为 1 且至少有一个相邻像素的值为 0 的像素。参考图11.14 中的 8 邻域表示，如果满足如下条件，则步骤 1 将一个轮廓点 p_1 标为要删除的点：

(a) $2 \leqslant N(p_1) \leqslant 6$

(b) $T(p_1) = 1$

(c) $p_2 \cdot p_4 \cdot p_6 = 0$

(d) $p_4 \cdot p_6 \cdot p_8 = 0$　　(11.1-4)

式中，$N(p_1)$是 p_1 的非零相邻像素数量，即

$$N(p_1) = p_2 + p_3 + \cdots + p_8 + p_9 \quad (11.1\text{-}5)$$

其中，p_i要么为 0，要么为 1，而 $T(p_1)$是排序序列 $p_2, p_3, \cdots, p_8, p_9, p_2$ 中从 0 到 1 转换的次数。例如，在图11.15中，有 $N(p_1) = 4$ 和 $T(p_1) = 3$。

p_9	p_2	p_3
p_8	p_1	p_4
p_7	p_6	p_5

图 11.14　细化算法所用的邻域像素排列

0	0	1
1	p_1	0
1	0	1

图 11.15　式(11.1-4)中条件(a)和(b)的说明。此时，$N(p_1) = 4$ 且 $T(p_1) = 3$

在步骤 2 中，条件(a)和(b)保持不变，但条件(c)和(d)变为

(c′) $p_2 \cdot p_4 \cdot p_8 = 0$

(d′) $p_2 \cdot p_6 \cdot p_8 = 0$　　(11.1-6)

834～835

把步骤 1 用于所考虑二值区域的每个边界像素。如果违反了条件(a)到(d)中的一个或多个，则所讨论点的值不变。如果满足所有的条件，则将该点标记为删除点，但该点要等到处理完所有边界点后才能删除。这种延迟可防止算法执行期间对数据结构的更改。在对所有的边界点应用步骤 1 后，删除已标记点(改为 0 值)。然后，按与步骤 1 完全相同的方式，对所得结果应用步骤 2。

这样，细化算法的一次迭代就由如下步骤组成：(1)应用步骤 1，标记待删除的边界点；(2)删除已标记点；(3)应用步骤 2 来标记剩余的待删除边界点；(4)删除已标记点。反复应用这个基本过程，直到再也没有待删除的点为止，此时算法终止，生成了该区域的骨架。

当轮廓点 p_1 仅有 1 个或 7 个值为 1 的邻域点时，就违反了条件(a)。仅有一个这样的邻点意味着 p_1 是一个骨架笔画的端点，显然不能删除这个点。如果 p_1 有 7 个这样的邻点，则删除 p_1 会对该区域产生腐蚀。当对宽度为 1 像素的笔画上的点应用这种方法时，就违反了条件(b)。因此，该条件在细化操作期间就防止了骨架线段的断裂。值的最小集合($p_4=0$ 或 $p_6=0$)或($p_2=0$ 和 $p_8=0$)同时满足条件(c)和(d)。这样，参考图 11.14 中的邻域排列，满足这些条件及条件(a)和(b)的点，就是边界中的一个东边界点或南边界点，或边界中西北角的点。无论哪种情况，p_1 都不是骨架的一部分，应被删除。类似地，下列值的最小集合($p_2=0$ 或 $p_8=0$)或($p_4=0$ 和 $p_6=0$)同时满足条件(c′)和(d′)，它们对应于北边界点或西边界点，或一个东南角点。注意，东北角点有 $p_2=0$ 和 $p_4=0$，既满足条件(c)和(d)，也满足条件(c′)和(d′)。对于西南角点有 $p_6=0$ 和 $p_8=0$，故也满足前述条件。

例 11.5　一个区域的骨架。

图11.16 显示了人腿骨的一幅分割图像，并叠加了该区域的骨架。在很大程度上，该骨架直观上看起来是正确的。在骨头的右“肩”处有两个分支，乍看之下，与左侧相对应，我们认为此处应只有一个分支。但要注意的是，(在长方向上)右肩要比左肩稍宽，这就是算法产生两个分支的原因。在骨架处理算法中，这种不可预测的行为很常见。

图 11.16　人腿骨和叠加在所示区域上的骨架

11.2　边界描绘子

本节我们考虑几种描绘区域边界的方法，在 11.3 节，我们重点关注区域描绘子。11.4 节和 11.5 节中的部分内容既适用于边界也适用于区域。

11.2.1　一些简单的描绘子

边界的长度是其最简单的描绘子之一。一条边界上的像素数量可以给出其长度的粗略近似。对于在两个方向上以单位间距定义的链码曲线，垂直分量和水平分量的数量加上对角分量的 $\sqrt{2}$ 倍，可给出曲线的准确长度。

边界 B 的直径定义为

$$\text{Diam}(B)=\max_{i,j}[D(p_i,p_j)] \tag{11.2-1}$$

式中，D 是一种距离测度(见 2.5.3 节)，p_i 和 p_j 是边界上的点。直径的值和连接组成该直径的两个端点的直线段(该直线称为边界的长轴)的方向是边界的有用描绘子。边界的短轴定义为与长轴垂直的直线。由边界与两个轴相交的 4 个外部点所组成的方框这样的长度[①]，可以完全包围该边界。刚刚描述的这个方框称为基本矩形，长轴与短轴之比称为边界的偏心率。偏心率也是一个有用的描绘子。

曲率定义为斜率的变化率。一般来说，获得数字边界上某个点处曲率的可靠测度较为困难，因为

836~837

① 不要将长轴和短轴与 11.4 节中定义的本征轴混淆。

这些边界往往都是局部“粗糙的”。然而，有时使用相邻边界线段的斜率差作为这两条线段交点处曲率的描绘子，被证明是有用的。例如，借用图11.6(c)所示边界的顶点本身也可以很好地描述曲率。在边界上顺时针方向移动时，如果点 p 处的斜率变化非负，则称顶点 p 是凸线段的一部分；否则，称 p 属于凹线段。使用斜率变化范围，可使某个点处的曲率更为精确。例如，如果斜率变化小于 10°，则认为 p 是一条近似直线线段的一部分；如果斜率变化超过 90°，则认为 p 是一个角点。必须小心使用这些描绘子，因为它们的解释取决于单条线段相对于整个边界的长度。

11.2.2 形状数

如 11.1.2 节中解释的那样，链码边界的一次差分取决于起始点。一条基于图11.3(a)所示四方向编码的边界的形状数，定义为最小量级的一次差分。形状数的阶 n 定义为其表示的数字个数。此外，对于闭合边界，n 为偶数，其值限制了不同形状的数量。图11.17显示了阶为 4, 6, 8 的所有形状，以及它们的链码表示、一次差分和相应的形状数。注意，一次差分是通过将链码作为循环序列来计算得到的，详见 11.1.2 节。尽管链码的一次差分与旋转无关，但总体而言，已编码的边界取决于网格的方向。归一化网格方向的一种方法是，使用前一节中定义的基本矩形的边对准链码网格。

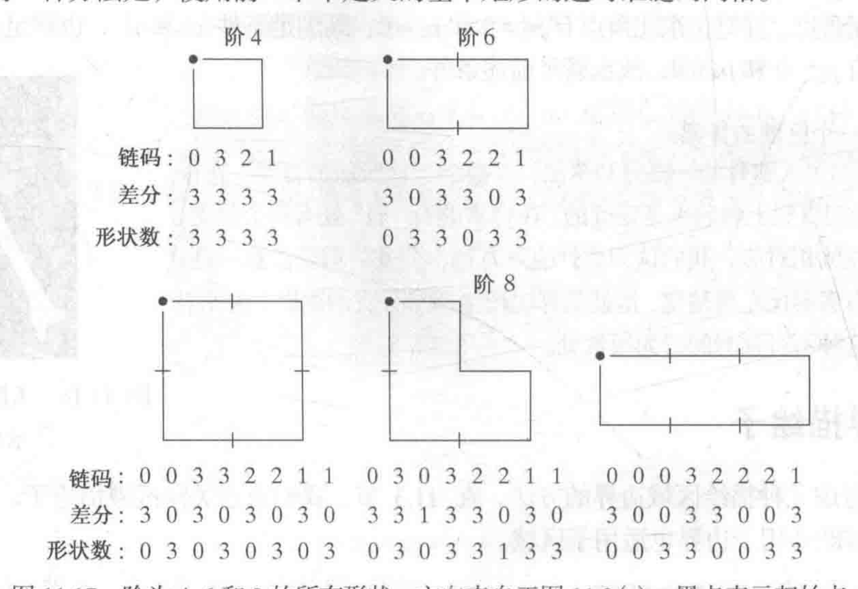


图 11.17 阶为 4, 6 和 8 的所有形状。方向来自于图 11.3(a)，黑点表示起始点

实际上，对于一个期望形状的阶，我们要找到阶为 n 的矩形，该矩形的偏心率(见前节的定义)最接近基本矩形，并使用这个新矩形来建立网格尺寸。例如，如果 $n = 12$，则阶为 12 的所有矩形(即周长为 12 的矩形)是 2×4, 3×3 和 1×5。如果 2×4 矩形的偏心率最匹配一个给定边界的基本矩形的偏心率，那么我们以这个基本矩形为中心建立一个 2×4 网格，并使用 11.1.2 节中给出的过程来得到链码。形状数是根据该编码的一次差分得出的。尽管选取网格间距的方式通常会使得形状数的阶等于 n，但低于该间距的边界有时会产生阶大于 n 的形状数。此时，我们指定阶数小于 n 的一个矩形，并且重复该过程，直到所得形状数的阶为 n。

例 11.6 计算形状数。

假设图11.18(a)中的边界指定的阶为 $n = 18$。为得到该阶数的形状数，我们需要按照刚才讨论的下列步骤来操作。第一步是找到基本矩形，如图11.18(b)所示。阶为 18 的最接近矩形为 3×6 矩形，这时要求按图 11.18(c)那样细分该基本矩形，其中链码方向与所得到的网格对齐。最后一步是得到链码，并使用该链码的一次差分来计算形状数，如图 11.18(d)所示。

837 ~ 839

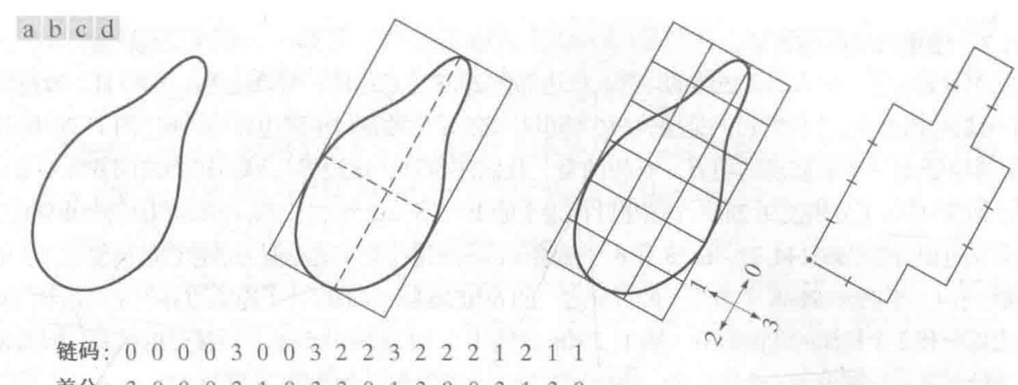


图 11.18　生成形状数的步骤

11.2.3　傅里叶描绘子

图11.19显示了 xy 平面内的一个 K 点数字边界。从任意点(x_0, y_0)开始，以逆时针方向在该边界上行进时，会遇到坐标对$(x_0, y_0), (x_1, y_1), (x_2, y_2), \cdots, (x_{K-1}, y_{K-1})$。这些坐标可以表示为 $x(k) = x_k,\ y(k) = y_k$ 的形式。使用这种表示法，边界本身可以表示为坐标序列 $s(k) = [x(k), y(k)],\ k = 0, 1, 2, \cdots, K-1$。此外，每个坐标对可当作一个复数来处理，即

$$s(k) = x(k) + \mathrm{j}y(k) \tag{11.2-2}$$

式中 $k = 0, 1, 2, \cdots, K-1$。也就是说，x 轴为复数序列的实轴，y 轴为复数序列的虚轴。尽管对该序列的解释是全新的，但边界本身的性质并未改变。当然，这种表示方法的一大优点是它将二维问题简化成了一维问题。

由式(4.4-6)可知，$s(k)$的离散傅里叶变换(DFT)为

$$a(u) = \sum_{k=0}^{K-1} s(k)\mathrm{e}^{-\mathrm{j}2\pi uk/K} \tag{11.2-3}$$

式中 $u = 0, 1, 2, \cdots, K-1$。复系数 $a(u)$称为边界的傅里叶描绘子。这些系数的傅里叶反变换可恢复 $s(k)$。也就是说，由式(4.4-7)有

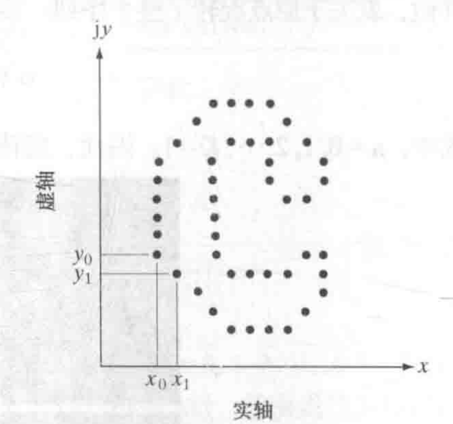


图 11.19　一条数字边界及其复序列表示。所示点(x_0, y_0)和点(x_1, y_1)是序列中(任意选择)的前两个点

$$s(k) = \frac{1}{K}\sum_{u=0}^{K-1} a(u)\mathrm{e}^{\mathrm{j}2\pi uk/K} \tag{11.2-4}$$

式中，$k = 0, 1, 2, \cdots, K-1$。然而，假设仅使用前P个傅里叶系数而不使用所有系数。这等同于在式(11.2-4)中令 $a(u) = 0,\ u > P-1$。结果为 $s(k)$的如下近似：

$$\hat{s}(k) = \frac{1}{K}\sum_{u=0}^{P-1} a(u)\mathrm{e}^{\mathrm{j}2\pi uk/P} \tag{11.2-5}$$

式中，$k = 0, 1, 2, \cdots, K-1$。尽管求 $\hat{s}(k)$ 的每个成分时仅使用了 P 项，但 k 的范围仍是从 0 到 $K-1$。也就是说，在近似边界中存在同样数量的点，但项数不像在每个点的重建中那样多。回顾第 4 章中关于傅里叶变换的讨论可知，高频成分说明精细细节，而低频成分决定全局形状。因此，P 越小，边界丢失的细节就越多，如下例所示。

840 ~ 841

例 11.7 使用傅里叶描绘子。

图 11.20(a) 显示了一个人的染色体的边界，该边界由 2868 个点组成。对该边界使用式(11.2-3)得到了相应的 2868 个傅里叶描绘子。本例的目的是检验减少傅里叶描绘子的数量对重建边界的影响。图 11.20(b) 显示了使用 2868 个描绘子的一半来重建的边界。有趣的是，我们注意到，该边界与原始边界之间并无明显的不同。图 11.20(c) 到(h) 显示了分别使用 2868 个傅里叶描绘子的 10%, 5%, 2.5%, 1.25%, 0.63%和 0.28%重建的边界。这些百分比分别近似等于 286, 144, 72, 36, 18 和 8 个描绘子，注意这些数字已近似为最接近的偶整数。这里的重点是 18 个描绘子（约为原来 2868 个描绘子的百分之一的 6/10）就有效地保持了原始边界的主要形状特征：4 个较长的突出部分和 2 个较深的弯曲部分。图 11.20(h) 是使用 8 个描绘子得到的不可接受的结果，因为丧失了主要特征。描绘子进一步减少到 4 个和 2 个，将分别导致一个椭圆和一个圆(见习题 11.13)。

如前例所示的那样，少数的傅里叶描绘子就可用于捕获边界的大体特征。这一性质很有用，因为这些系数携带有形状信息。因此，这些描绘子可用做区分不同边界形状的基础，正如我们在第12章中讨论的那样。

我们已多次声明，描绘子应尽可能对平移、旋转和尺度变化不敏感。此时，结果将取决于点被处理的顺序，故另一个约束条件是，描绘子对起始点应不敏感。傅里叶描绘子并非直接对这些几何变化不敏感，但这些参数的变化可能与描绘子的简单变换是相关的。例如，考虑旋转变换，且回忆可知，基本数学分析中关于复平面的原点将一个点旋转θ角是通过乘以因子 $e^{j\theta}$ 来实现的。对 $S(k)$的每个点这样做，就关于原点旋转了整个序列。旋转后的序列为 $s(k)e^{j\theta}$ ，其傅里叶描绘子为

$$a_r(u) = \sum_{k=0}^{K-1} s(k)e^{j\theta}e^{-j2\pi uk/K} = a(u)e^{j\theta} \tag{11.2-6}$$

式中，$u = 0, 1, 2, \cdots, K-1$。因此，旋转变换仅通过乘以一个常数项 $e^{j\theta}$ ，就等同地影响了所有的系数。

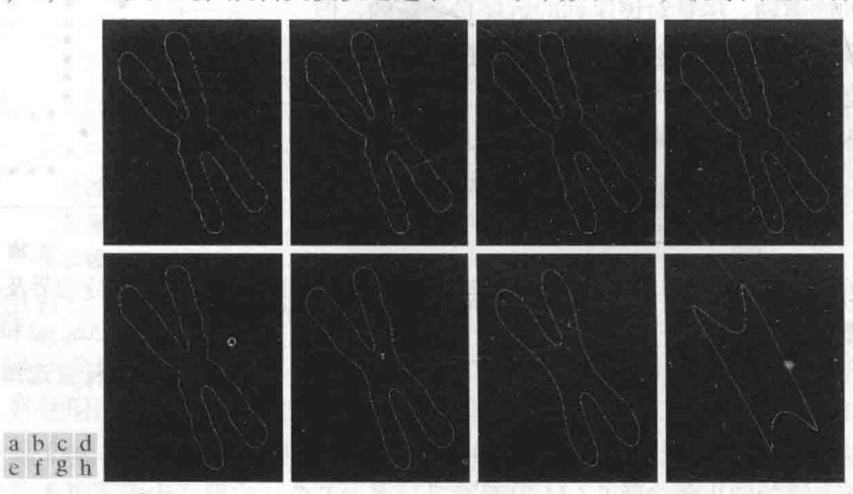


图 11.20 (a) 人的染色体的边界(2868 个点)；(b)～(h) 分别使用 1434, 286, 144, 72, 36, 18 和 8 个傅里叶描绘子重建的边界。这些数字分别近似为 2868 个点的 50%, 10%, 5%, 2.5%, 1.25%, 0.63%和 0.28%

表11.1 总结了经历旋转、平移、尺度和起始点变化的边界序列 $S(k)$的傅里叶描绘子。符号 Δ_{xy} 定义为 $\Delta_{xy} = \Delta x + j\Delta y$ ，所以符号 $s_t(k) = s(k) + \Delta_{xy}$ 将该序列重定义(平移)为

$$s_t(k) = [x(k) + \Delta x] + j[y(k) + \Delta y] \tag{11.2-7}$$

换句话说，平移是对边界上的所有坐标添加一个常数位移。注意，除了 $u = 0$ 时变为冲激函数 $\delta(u)$[①]

841
～
843

① 回忆第 4 章可知，常数的傅里叶变换是一个位于原点的冲激，且该冲激处处为零。

外，平移变换对描绘子无影响。最后，表达式 $s_p(k)=s(k-k_0)$意味着将序列重新定义为

$$s_p = x(k-k_0) + \mathrm{j}y(k-k_0) \tag{11.2-8}$$

它仅将序列的起始点从 $k=0$ 更改为 $k=k_0$。表 11.1 中的最后一项表明，起始点的变化会以不同(但已知)的方式影响到所有描绘子，在这一意义上说，乘项 $a(u)$取决于 u。

表 11.1　傅里叶描绘子的一些基本性质

变　换	边　界	傅里叶描绘子
恒等	$s(k)$	$a(u)$
旋转	$sr(k)=s(k)\mathrm{e}^{\mathrm{j}\theta}$	$ar(u)=a(u)\mathrm{e}^{\mathrm{j}\theta}$
平移	$st(k)=s(k)+\Delta_{xy}$	$at(u)=a(u)+\Delta_{xy}\delta(u)$
缩放	$ss(k)=\alpha s(k)$	$as(u)=\alpha a(u)$
起始点	$sp(k)=s(k-k_0)$	$ap(u)=a(u)\,\mathrm{e}^{-\mathrm{j}2\pi k_0 u/K}$

11.2.4　统计矩

边界线段(和标记图波形)的形状可使用统计矩来定量描述，如均值、方差和高阶矩。为了解如何实现这种定量描述，考虑图11.21(a)，它显示了一段边界，而图11.21(b)显示了以任意变量 r 的一维函数 $g(r)$描述的线段。该函数是这样获得的：先将该线段的两个端点连接起来，然后旋转该直线线段，直至其为水平线段。此时，所有点的坐标也旋转相同的角度。

关于概率论的简要回顾，请查阅本书的网站。

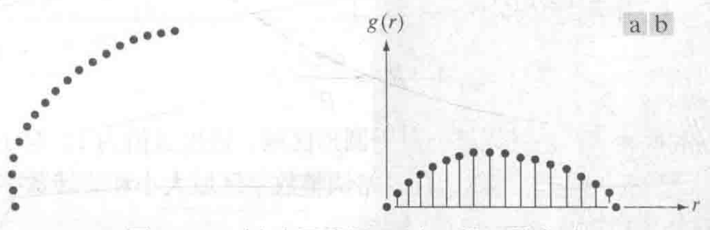


图 11.21　(a)边界线段；(b)一维函数表示

让我们将幅度 g 视为一个离散随机变量 v，并形成一个幅度直方图 $p(v_i)$, $i=0, 1, 2,\cdots, A-1$，其中 A 是我们分割幅度尺度的离散幅度增量数。然后，记住 $p(v_i)$是值 v_i 出现的概率估计，故由式(3.3-17)可知关于其均值的 v 的第 n 阶矩为

$$\mu_n(v) = \sum_{i=0}^{A-1}(v_i - m)^n p(v_i) \tag{11.2-9}$$

式中，

$$m = \sum_{i=0}^{A-1} v_i p(v_i) \tag{11.2-10}$$

量 m 为 v 的均值或平均值，μ_2 为 v 的方差。通常，仅使用前几个矩通常不能区分明显不同形状的信号。

一种替代方法是将 $g(r)$归一化为单位面积，并把它当作直方图来处理。换句话说，$g(r_i)$现在作为值 r_i 出现的概率来处理。此时，r 为一个随机变量，故矩为

$$\mu_n(r) = \sum_{i=0}^{K-1}(r_i - m)^n g(r_i) \tag{11.2-11}$$

式中，

$$m = \sum_{i=0}^{K-1} r_i g(r_i) \tag{11.2-12}$$

843～844

式中，K 是边界上的点数，$\mu_n(r)$直接与 $g(r)$的形状相关。例如，二阶矩 $\mu_2(r)$度量曲线关于 r 的均值的扩展程度，而三阶矩 $\mu_3(r)$度量曲线关于均值的对称性。

基本上，我们完成的工作是将任务的描述简化为一维函数的描述。尽管到目前为止矩是最流行的方法，但它们不是用于此目的的唯一描绘子。例如，另一种方法是计算一维离散傅里叶变换，得到其频谱，并使用频谱的前 q 个成分来描述 $g(r)$。与其他技术相比，矩方法的优点是矩的实现非常简单，且携带了边界形状的“物理”解释。图11.21清楚地表明了这种方法对旋转的不敏感性。需要时，通过缩放 g 值和 r 值的范围，可实现尺寸的归一化。

11.3 区域描绘子

本节考虑几种描述图像区域的方法。记住，使用边界和区域相结合的描绘子在实践中是很普遍的。

11.3.1 一些简单的描绘子

区域的面积定义为该区域中像素的数量。区域的周长是其边界的长度。尽管面积和周长有时也被用做描绘子，但它们主要应用于感兴趣区域尺寸不变的情形。这两个描绘子频繁用于度量一个区域的致密性，致密性定义为(周长)^2/面积。另一个稍微不同的(标量乘子内)致密性描绘子是圆度率，即一个区域的面积与具有相同周长的一个圆(最致密形状)的面积之比。周长为 P 的一个圆的面积为 $P^2/4\pi$。因此，圆度率 R_c 由下式给出：

$$R_c = \frac{4\pi A}{P^2} \tag{11.3-1}$$

式中，A 是所讨论区域的面积，P 是其周长。对于圆形区域，该度量值为 1，对于方形区域，该度量值为$\pi/4$。致密性是一个无量纲的测度，当然，在忽略调整数字区域大小和旋转数字区域时可能引入的计算误差的情形下，它对均匀尺度变化和方向不敏感。

用做区域描绘子的其他简单测度包括灰度级的均值和中值、最小灰度值和最大灰度值，以及其值高于和低于均值的像素数。

例 11.8 使用面积计算从图像中提取信息。

即使是像归一化面积这样的简单区域描绘子，对于从图像中提取信息也是相当有用的。例如，图 11.22 显示了美洲的卫星红外图像。如 1.3.4 节中讨论的那样，这些图像提供了人类定居地的整体数量。用于收集这些图像的传感器能够探测可见光和近红外发射光，如灯光、火光和闪光。图像旁边的表格(按从上到下的区域)给出了所有 4 个区域中被白色(灯光)占据的面积与所有发光面积之比。例设这种简单测度可用区域消耗的电能给出相对估计。通过关于每个区域的陆地面积、人口数等将其归一化，可进一步细化数据。

11.3.2 拓扑描绘子

拓扑特性对于图像平面区域的整体描述是很有用的。简单来说，拓扑学研究未受任何变形影响的图形的性质，前提是该图形未被撕裂或粘连(有时称为橡皮膜变形)。例如，图11.23显示了一个带有两个孔洞的区域。如果一个拓扑描绘子由该区域内的孔洞数量来定义，那么这种性质明显不受拉伸或旋转变换的影响。然而，一般来说，如果该区域被撕裂或折叠，那么孔洞数会发生变化。注意，由于拉伸会影响距离，故拓扑特性与距离或基于距离测度概念的任何特性无关。

844 ~ 845

另一个对区域描述有用的拓扑特性是连通分量的数量。2.5.2 节定义了区域的连通分量。图 11.24 显示了一个具有 3 个连通分量的区域(关于计算连通分量的算法，请参阅 9.5.3 节)。



区域编号（从上至下）	每个区域的灯光与总灯光之比
1	0.204
2	0.640
3	0.049
4	0.107



图 11.22　夜间美洲的红外图像(图像由 NOAA 提供)

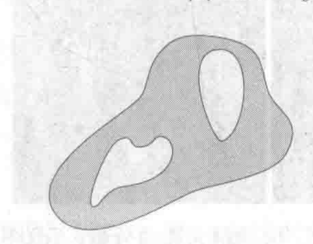




图 11.23　带有两个孔洞的一个区域

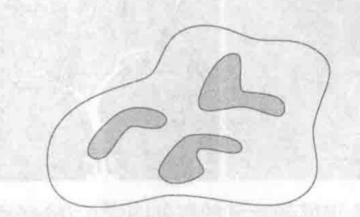


图 11.24　带有 3 个连通分量的一个区域

图形中孔洞的数量 H 和连通分量的数量 C，可用于定义欧拉数 E：

$$E = C - H \tag{11.3-2}$$

欧拉数也是一种拓扑特性。例如，图11.25 中所示的区域有分别等于 0 和 – 1 的欧拉数，因为“A”有一个连通分量和一个孔洞，而“B”有一个连通分量和两个孔洞。

使用欧拉数，可非常简单地解释由直线段表示的区域(称为多边形网络)。图11.26 显示了一个多边形网络。将这样一个网络的内部区域分类为面和孔通常很重要。用 V 表示顶点数，用 Q 表示边数，用 F 表示面数，那么可得出称为欧拉公式的如下关系：

845～847

$$V - Q + F = C - H$$

由式(11.3-2)可以看出上式等于欧拉数：

$$V-Q+F=C-H=E \tag{11.3-3}$$

图11.26中的网络有7个顶点、11条边、2个面、1个连通区域和3个孔，因此欧拉数为–2：

$$7-11+2=1-3=-2$$

拓扑描绘子提供了一个附加特征，该特征在表征某个场景中的区域时通常很有用。

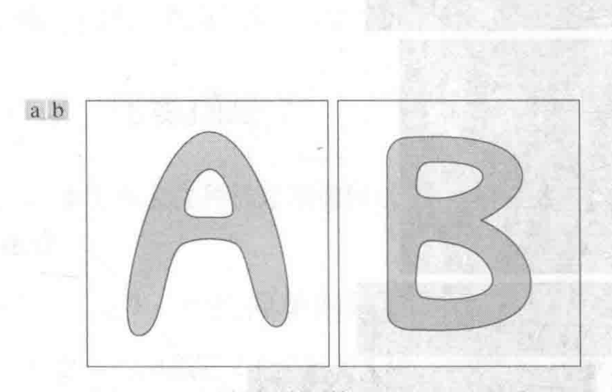


图 11.25 欧拉数分别等于 0 和–1 的区域

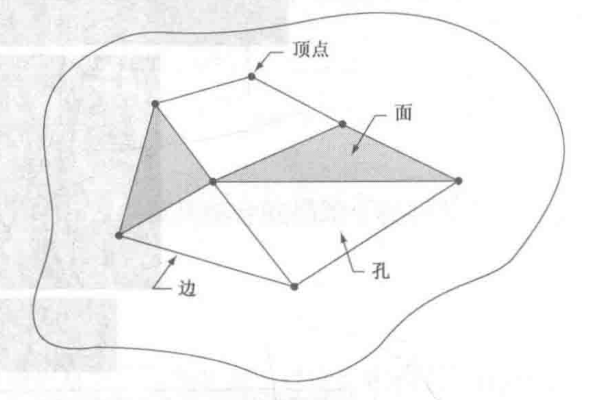


图 11.26 包含一个多边形网络的区域

例 11.9 使用连通分量在已分割图像中提取最大特征。

图 11.27(a)显示了美国华盛顿特区的一幅大小为 512×512 的 8 比特图像，该图像是由 NASA 的地球资源(LANDSAT)卫星拍摄的。这幅特殊的图像是近红外波段图像(详见图 1.10)。假设我们只使用这幅图像来分割河流(与使用多光谱图像相反，它可简化任务)。由于河流在图像中是相当暗的均匀区域，因此阈值处理方法显然值得一试。在使河流变为一个不连通区域之前，用可能的最高阈值为图像设置门限所得的结果如图 11.27(b)所示。阈值是人为选择的，这一点说明，在这种情况下，没有图像的其他区域出现在阈值化后的结果中，仅依靠河流本身是不可能分割河流的。本例的目的是说明如何使用连通分量来“完成”分割。

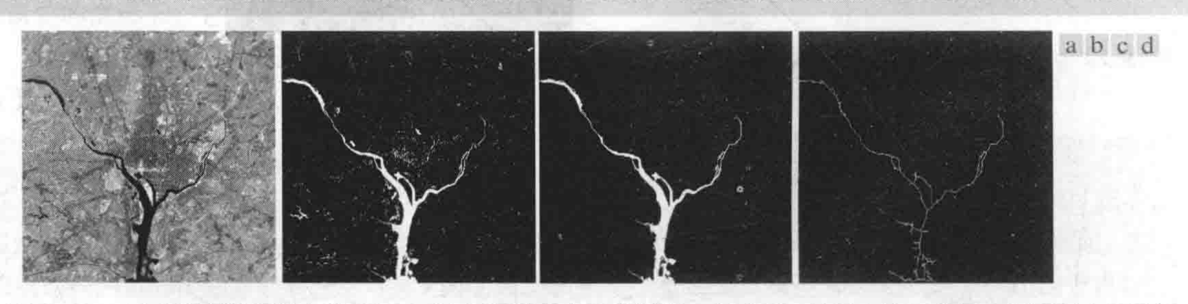


图 11.27 (a)华盛顿特区的红外图像；(b)经阈值处理后的图像；(c)图(b)的最大连通分量；(d)图(c)的骨架

图 11.27(b)中的图像有 1591 个连通分量(使用 8 连通得到)，其欧拉数为 1552，由此推出孔洞数为 39。图 11.27(c)显示了具有最大元素数量(8479)的连通分量。这是我们期望的结果，因为我们已知使用一个阈值不能从图像中由河流本身来分割河流。注意，所得结果非常清晰。如果我们想要度量河流的每个支流的长度，那么可以使用该连通分量的骨架[见图11.27(d)]。换句话说，骨架中每个分支的长度是它所表示的河流分支长度的合理近似。

11.3.3 纹理

847~849

描绘区域的一种重要方法是量化该区域的纹理内容。尽管不存在纹理的正式定义，但在直觉上，这种描绘子提供了诸如平滑度、粗糙度和规律性等特性的测度(图11.28中给出了一些例子)。图像处理

中用于描述区域纹理的三种主要方法是统计法、结构法和频谱法。统计法生成诸如平滑、粗糙、粒状等纹理特征。结构法处理图像像元的排列，如基于规则间距平行线的纹理描述。频谱法基于傅里叶频谱的特性，主要用于检测图像中的全局周期性，方法是识别频谱中的高能量的窄波峰。

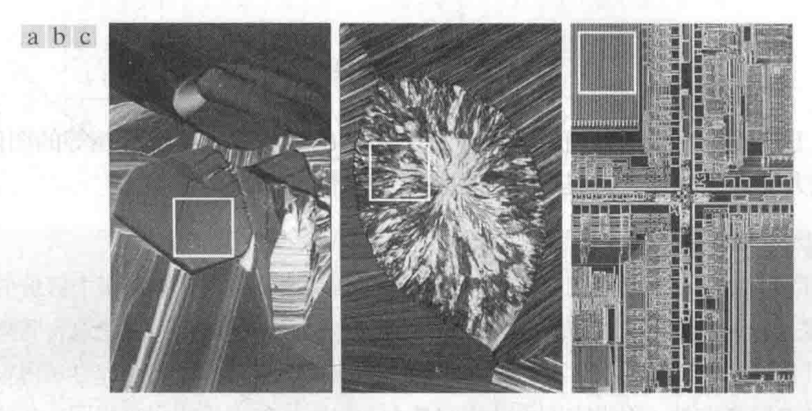


图 11.28 白色方块从左到右表示平滑纹理、粗糙纹理和规则纹理。图像从左到右分别是一块超导体、人类胆固醇和一个微处理器的光学显微镜图像(图像由佛罗里达州立大学的 Michael W. Davidson 博士提供)

统计法

描述纹理的最简方法之一是使用一幅图像或一个区域的灰度级直方图的统计矩。令 z 是表示灰度的一个随机变量，并令 $p(z_i), i = 0, 1, 2, \cdots, L-1$ 为相应的直方图，其中 L 是不同灰度级的数量。由式(3.3-17)可知，关于其均值的 z 的第 n 阶矩为

$$\mu_n(z) = \sum_{i=0}^{L-1}(z_i - m)^n p(z_i) \tag{11.3-4}$$

式中，m 是 z 的均值(平均灰度)：

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \tag{11.3-5}$$

注意，由式(11.3-4)可知$\mu_0 = 1$ 和$\mu_1 = 0$。二阶矩［方差$\sigma^2(z) = \mu_2(z)$］在纹理描述中特别重要。它是灰度对比度的测度，可用于建立相对平滑度的描绘子。例如，测度

$$R(z) = 1 - \frac{1}{1 + \sigma^2(z)} \tag{11.3-6}$$

对于恒定灰度区域为 0(该区域的方差为零)，而对于较大的 $\sigma^2(z)$值，其接近于 1。因为对于灰度级图像，方差值是随灰度值增大而增大的，例如，对于值域为[0, 255]的灰度级图像，最好将方差归一化到区间[0, 1]内，以便于在式(11.3-6)中使用。在式(11.3-6)中将 $\sigma^2(z)$ 除以$(L-1)^2$，可实现这种归一化。标准差$\sigma(z)$ 也经常用做纹理的一种测度，因为标准差的值对于很多人来说更为直观。

三阶矩

$$\mu_3(z) = \sum_{i=0}^{L-1}(z_i - m)^3 p(z_i) \tag{11.3-7}$$

是直方图偏斜度的测度，而四阶矩是直方图相对平坦度的测度。五阶矩和更高阶矩不容易与直方图形状联系起来，但它们的确提供了纹理内容的进一步分辨。另外一些有用的、基于直方图的纹理测度，包括由下式给出的“一致性”测度：

849 ~ 851

$$U(z)=\sum_{i=0}^{L-1}p^2(z_i) \tag{11.3-8}$$

和平均熵测度：

$$e(z)=-\sum_{i=0}^{L-1}p(z_i)\log_2 p(z_i) \tag{11.3-9}$$

因为 p 在区间[0, 1]内取值，且这些值的和为 1，故测度 U 对所有灰度级都相等的图像有最大值(极大一致性)，并从最大值开始降低。熵是可变性的测度，对恒定图像其为零。

例 11.10　基于直方图的纹理测度。

表 11.2 总结了前述图 11.28 中强调的三种纹理的测度值。均值仅告诉我们每个区域的平均灰度，它仅用做灰度的大致概念，不是真正的纹理。标准差提供的信息更为丰富；数字清楚地表明第一个纹理的灰度级变化明显小于其他两个纹理(即第一个纹理更平滑)。用这种测度时，粗糙纹理显示得很清楚。如期望的那样，该结论对于 R 同样有效，因为它度量的内容基本上与标准差度量的内容相同。三阶矩通常用于确定直方图的对称程度及它们是向左偏斜(负值)还是向右偏斜(正值)。这给出了灰度级是偏向均值的暗端还是偏向亮端的粗略概念。就纹理而论，仅当测度间的变化较大时，由三阶矩导出的信息才是有用的。研究一致性测度后，我们再次得出结论：第一幅子图像更为平滑(比其他子图像更均匀)，且最大随机(最低一致性)对应于粗糙纹理。这并不让人感到意外。最后，熵值变化与一致性测度相反，它与一致性测度的结论相同。第一幅子图像的灰度级变化最小，而粗糙图像的灰度级变化最大。规则纹理的灰度级变化介于前述两种测度之间。

表 11.2　图 11.28 中所示子图像的纹理测度

纹　理	均　值	标准差	(归一化后的)R	三阶矩	一致性	熵
平滑	82.64	11.79	0.002	–0.105	0.026	5.434
粗糙	143.56	74.63	0.079	–0.151	0.005	7.783
规则	99.72	33.73	0.017	0.750	0.013	6.674

仅使用直方图计算得到的纹理测度不携带像素彼此之间的相对位置的信息。在描述纹理时，这很重要，并且将这种类型的信息合并到纹理分析过程中的一种方法是，不仅要考虑灰度的分布，还要考虑图像中像素的相对位置。

令 Q 是定义两个像素彼此相对位置的一个算子，并考虑一幅具有 L 个可能灰度级的图像 f。令 $\boldsymbol{G}$ 为一个矩阵，其元素 g_{ij} 是灰度为 z_i 和 z_j 的像素对出现在 f 中由 Q 指定位置处的次数，其中 $1 \leqslant i, j \leqslant L$。按这种方法形成的矩阵称为*灰度级(或灰度)共生矩阵*。当含义明确时，$\boldsymbol{G}$ 简单地称为一个*共生矩阵*。

注意，我们正使用灰度范围[1, L]替代我们常用范围[0, $L-1$]。这样做的目的是让灰度值对应于“传统的”矩阵索引(即灰度值 1 对应于 $\boldsymbol{G}$ 的第一个行列索引)。

图 11.29 显示了一个如何用 $L=8$ 和由“直接面对右边的一个像素”(一个像素的相邻像素定义为紧靠右边的像素)定义的位置算子 Q 构造共生矩阵的例子。左侧的阵列是所考虑的小图像，右侧的阵列是矩阵 $\boldsymbol{G}$。我们看到，$\boldsymbol{G}$ 的元素(1, 1)是 1，因为在 f 中，值为 1 的像素的右侧值也为 1 的像素仅出现了 1 次。类似地，$\boldsymbol{G}$ 的元素(6, 2)是 3，因为在 f 中，值为 6 的像素的右侧值为 2 的像素出现了 3 次。按这种方式，可计算出 $\boldsymbol{G}$ 的其他元素。例如，如果我们已将 Q 定义为“右边一个像素和上面一个像素”，则 $\boldsymbol{G}$ 中的位置(1, 1)会是 0，因为在 f 内由 Q 指定的位置处不存在右边一个 1 和上面一个 1 的情形。另一方面，$\boldsymbol{G}$ 中的位置(1, 3), (1, 5)和(1, 7)都是 1，因为在 f 内由 Q 定义的位置处，右侧像素分别为 3, 5 和 7 的 1 值像素都出现了 1 次。作为练习，请你使用 Q 的这个定义计算 $\boldsymbol{G}$ 的所有元素。

851 ~ 853

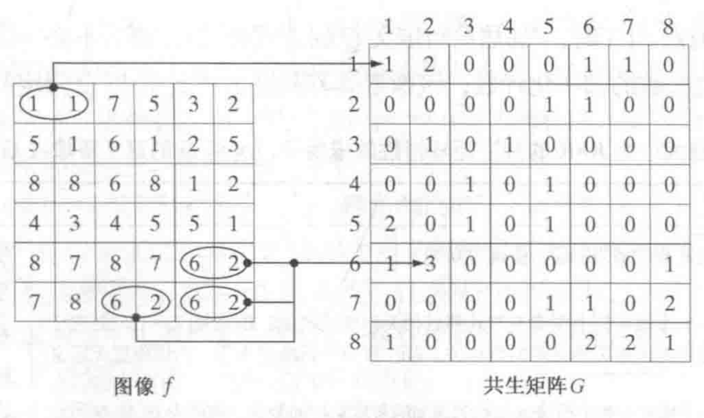


图 11.29　生成共生矩阵的方式

图像中可能的灰度级数决定了矩阵 $\boldsymbol{G}$ 的大小。对于一幅 8 比特图像(256 个可能的灰度级)，$\boldsymbol{G}$ 的大小为 256×256。使用一个矩阵时，这不成问题，但如例 11.11 中所示，有时会使用共生矩阵序列。为了减少计算负担，经常使用的一种方法是，将灰度量化为几段，以便于管理矩阵 $\boldsymbol{G}$ 的大小。例如，在 256 级灰度的情形下，通过令前 32 个灰度级等于 1，令接下来的 32 个灰度级等于 2，以此类推，可完成这种量化工作。这将得到一个大小为 8×8 的矩阵。

满足 Q 的像素对的总数 n，等于 $\boldsymbol{G}$ 的元素之和(在上例中 $n = 30$)。因此，

$$p_{ij} = g_{ij} / n$$

是满足 Q 的一个值为(z_i, z_j)的点对的概率估计。这些概率的值域为[0, 1]，且它们的和为 1：

$$\sum_{i=1}^{K}\sum_{j=1}^{K} p_{ij} = 1$$

式中，K 是方阵 $\boldsymbol{G}$ 的行(或列)数。

因为 $\boldsymbol{G}$ 取决于 Q，所以选择一个合适的位置算子并分析 $\boldsymbol{G}$ 的元素，可以检测灰度纹理模式的存在。表 11.3 中列出了一组表征 $\boldsymbol{G}$ 的内容的有用描绘子。(表中第二行的)相关描绘子中使用的量定义如下：

$$m_r = \sum_{i=1}^{K} i \sum_{j=1}^{K} p_{ij}, \qquad m_c = \sum_{j=1}^{K} j \sum_{i=1}^{K} p_{ij}$$

和

$$\sigma_r^2 = \sum_{i=1}^{K} (i - m_r)^2 \sum_{j=1}^{K} p_{ij}, \qquad \sigma_c^2 = \sum_{j=1}^{K} (j - m_c)^2 \sum_{i=1}^{K} p_{ij}$$

如果我们令

$$P(i) = \sum_{j=1}^{K} p_{ij} \quad 和 \quad P(j) = \sum_{i=1}^{K} p_{ij}$$

则前述公式可以写为

$$m_r = \sum_{i=1}^{K} iP(i), \qquad m_c = \sum_{j=1}^{K} jP(j), \qquad \sigma_r^2 = \sum_{i=1}^{K} (i - m_r)^2 P(i), \qquad \sigma_c^2 = \sum_{j=1}^{K} (j - m_c)^2 P(j)$$

参考式(11.3-4)、式(11.3-5)和它们的解释，我们看到 m_r 是沿归一化后的 $\boldsymbol{G}$ 的行计算的均值，m_c 是沿着 $\boldsymbol{G}$ 的列计算的均值。类似地，σ_r 和 σ_c 是分别沿行和列计算的标准差(方差的平方根)。这些项都是标量，与 $\boldsymbol{G}$ 的大小无关。

853 ~ 855

记住，当我们研究表 11.3 时，“邻居”与定义 Q 的方式有关(即邻居不必一定是邻近的)，且 p_{ij} 只不过是在 f 内像素发生次数的归一化计数，该像素具有灰度 z_i 和 z_j 并与 Q 指定的位置有关。

表 11.3 用于表征大小为 $K\times K$ 的共生矩阵特性的描绘子，p_{ij} 是 G 的第 ij 项除以 G 的元素总和的数

描 绘 子	解 释	公 式
最大概率	度量 **G** 的最强响应，值域为[0, 1]	$\max_{ij}(p_{ij})$
相关	一个像素在整个图像上与其邻居相关程度的测度。值域是[1, −1]，对应于完美的正相关和完美的负相关。若任意一个标准差为零，则该测度无定义	$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r\sigma_c}$ $\sigma_r \neq 0; \sigma_c \neq 0$
对比度	一个像素在整个图像上与其邻居间的灰度对比的测度。值域为 0(当 **G** 恒定时)到$(K-1)^2$	$\sum_{i=1}^{K}\sum_{j=1}^{K}(i-j)^2 p_{ij}$
一致性(也称为能量)	值域为[0, 1]的一致性测度。对于恒定图像，一致性为 1	$\sum_{i=1}^{K}\sum_{j=1}^{K}p_{ij}^2$
同质性	**G** 中元素对角线分布的空间紧密度的测度。值域为[0, 1]，当 **G** 是对角矩阵时，其值最大	$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{p_{ij}}{1+\|i-j\|}$
熵	**G** 中元素随机性的测度。当所有 p_{ij} 项为零时熵为 0，当所有 p_{ij} 项相等时熵最大。最大值是 $2\log_2 K$ [见关于熵的公式(11.3-9)]	$-\sum_{i=1}^{K}\sum_{j=1}^{K}p_{ij}\log_2 p_{ij}$

例 11.11 使用描绘子表征共生矩阵。

图 11.30(a)到(c)显示了分别由随机模式、水平周期(正弦)模式和混合像素模式组成的图像。本例有两个目的：(1)对(从上到下)对应于这些图像的 3 个共生矩阵 $\mathbf{G}_1$, $\mathbf{G}_2$ 和 $\mathbf{G}_3$，说明表 11.3 中描绘子的值，(2)阐明如何使用共生矩阵序列来检测图像中的纹理模式。

图 11.31 以图像形式显示了共生矩阵 $\mathbf{G}_1$, $\mathbf{G}_2$ 和 $\mathbf{G}_3$。这些矩阵是用 $L = 256$ 和位置算子“右侧紧邻一个像素”得到的。在这些图像中，坐标(i, j)处的值是 f 内具有灰度 z_i 和 z_j 的像素对在由 Q 指定的位置处出现的次数，因此图 11.31(a)是一幅随机图像并不奇怪。

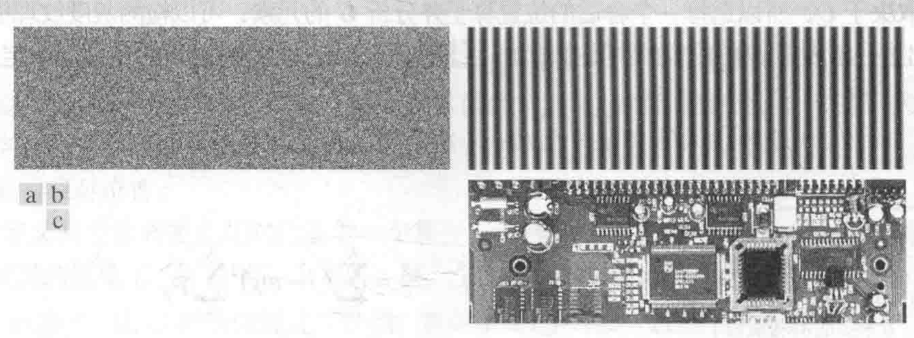


图 11.30 其像素具有(a)随机模式、(b)周期模式和(c)混合纹理模式的图像。每幅图像的大小都为 263×800 像素

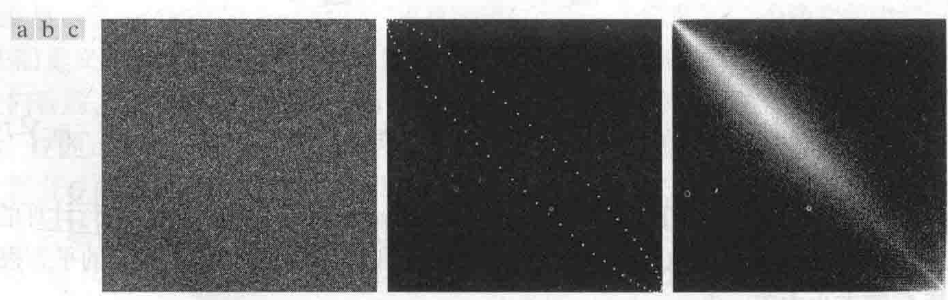


图 11.31 大小为 256×256 的共生矩阵 $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$，从左到右对应于图 11.30 中的图像

855 ~ 856

图11.31(b)更有趣。第一个明显的特征是关于主对角线对称。由于正弦波的对称性，像素对(z_i, z_j)的计数与像素对(z_j, z_i)的计数相同，从而产生了一个对称的共生矩阵。$\boldsymbol{G}_2$ 的非零元素是稀疏的，因为水平正弦波中水平相邻像素间的差值相对较小。数字化后的正弦波是阶梯状的，每个阶梯的高度和宽度取决于表示函数中使用的频率和幅度级数，这些有助于解释这些概念。

图 11.31(c)中共生矩阵 $\boldsymbol{G}_3$ 的结构更复杂。沿主对角线也聚集了高计数值，但它们的分布与 $\boldsymbol{G}_2$ 相比更密集，这一特征表明图像的灰度值有着丰富的变化，但相邻像素之间的灰度很少有较大的跳变。观察图 11.31(c)，我们看到图像中存在几个由较小灰度变化表征的较大区域。灰度的高跳变出现在物体的边界处，但与整个较大区域上的中等灰度跳变相比，这些计数较小，因此，如我们在第 3 章讨论的那样，图像显示器同时显示高值和低值的能力使得它们不太明显。

前面的观察是定性的。为定量化共生矩阵的内容，我们需要表11.3 中的那些描绘子。表11.4 给出了为图 11.31 中的 3 个共生矩阵计算的这些描绘子的值。注意，如之前讨论的那样，为使用这些描绘子，共生矩阵必须用它们的元素之和除它们来归一化。通过查看图11.30 中的图像和图 11.31 中对应的共生矩阵，表 11.4 中的各项与我们期望的结果一致。例如，考虑表 11.4 中的“最大概率”一列。最大概率对应于第三个共生矩阵，这告诉我们与其他两个矩阵相比，这个矩阵有最大数量的计数(像素对在图像内由 Q 指定的位置处出现的最大次数)。这与前面我们对 $\boldsymbol{G}_3$ 的分析一致。第二列指出最高相关对应于 $\boldsymbol{G}_2$，从而告诉我们第二幅图像中的灰度是高度相关的。图 11.30(b)中反复出现的正弦模式揭示了为什么会这样的原因。注意，$\boldsymbol{G}_1$ 的相关基本为零，实际上这表明相邻像素不相关，这是随机图像的一个特性，如图 11.30(a)中的图像。

对比度描绘子对于 $\boldsymbol{G}_1$ 最高，对于 $\boldsymbol{G}_2$ 最低。从而我们看到，图像的随机性越低，其对比度往往也越低。通过研究图 11.31 所示的矩阵，我们就会明白其原因。$(i-j)^2$ 项是 $1 \leqslant i, j \leqslant L$ 的整数差，故它们对于任何 $\boldsymbol{G}$ 都是相同的。因此，归一化共生矩阵中元素的概率就成为决定对比度值的因子。虽然 $\boldsymbol{G}_1$ 有最小的最大概率，但其他两个矩阵有许多零和接近零的概率(图 11.31 中的暗色区域)。记住，值 $\boldsymbol{G}/n$ 之和为 1，故很容易了解对比度描绘子随着随机性的增加而增大的原因。

使用类似的方式，可以解释剩下的三个描绘子。一致性会随概率平方值的增大而增加。这样，图像中的随机性越低，一致性描绘子就越高，如表 11.4 中的第五列所示。同质性度量 $\boldsymbol{G}$ 值关于主对角线的集中度。分母项$(1+|i-j|)$对所有三个共生矩阵是相同的，且分母项的值在 i 和 j 的值接近(即接近主对角线)时会减小。这样，靠近主对角线具有高概率值(分子项)的矩阵将有最高的同质性值。如前面讨论的那样，这样一个矩阵将对应于具有“丰富的”灰度级内容的图像和灰度值较慢变化的区域。表 11.4 中第六列的各项与这种解释是一致的。

表中最后一列中的各项是共生矩阵中的随机性测度，这些测度依次为对应图像中的随机性测度。如期望的那样，$\boldsymbol{G}_1$ 有最高的值，因为其源图像完全是随机的。其他两项的解释不言自明。注意，对 $\boldsymbol{G}_1$ 的熵的测度接近理论最大值 16($2\log_2 256 = 16$)。图 11.30(a)中的图像由均匀噪声组成，所以每个灰度级都有一个近似相等的出现概率，这是表 11.3 中为最大熵声明的一个条件。

至此，我们已处理了每幅图像及它们的共生矩阵。(在不查看图像的情形下)假设我们想要“发现”这些图像中是否存在任何包含重复分量(即周期纹理)的部分。完成该任务的一种方法是，增大邻点间的距离，由这些图像导出共生矩阵，并检查共生矩阵序列的相关描绘子。如前面提及的那样，在处理共生矩阵序列时，为减小矩阵的大小并减轻相应的计算负担，通常要量化灰度数。下边的结果是使用 $L=8$ 得到的。

图 11.32 显示了作为水平“偏移”(即相邻像素间的水平距离)从 1(对相邻像素)到 50 时的函数，相关描绘子的曲线。图 11.32(a)表明所有的相关值接近 0，这说明在随机图像中未找到这样的模式。图 11.32(b)中所示相关的形状清楚地表明，输入图像在水平方向是正弦的。注意，相关函数以一个高值开始，然后随着相邻像素间的距离的增大而降低，再后重复自身。

856 ~ 858

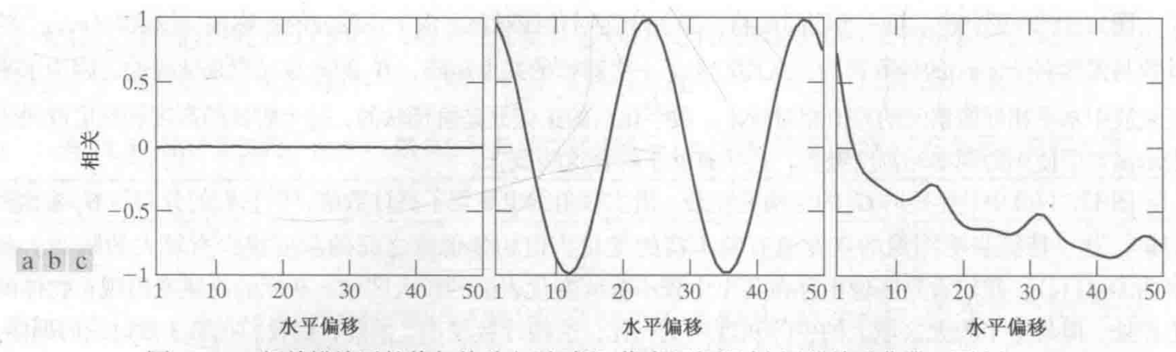


图 11.32　相关描绘子的值与偏移("相邻"像素间的距离)间的关系曲线，图(a)、图(b)和图(c)分别对应于图 11.30 中的噪声图像、正弦图像和电路板图像

图 11.32(c)表明，与电路板图像相关联的相关描绘子开始时下降，但在偏移距离为 16 像素时有一个较强的峰值。分析图 11.30(c)所示图像表明，顶部的焊接点形成了间距约为16像素的重复模式(见图11.33)，下一个主要的峰值出现在偏移距离 32 处，它是由同一模式引起的，但该波峰的幅度较低，因为在该偏移距离处的重复次数小于偏移距离 16 像素处的重复次数。类似的观察可以解释偏移距离 48 像素处的更小峰值。

图像中还存在其他重复的模式，但从 256 灰度级到 8 灰度级的粗糙量化已使它们变得模糊。

结构法

如本节开始时提到的那样，第二类纹理描述是基于结构概念的。假设我们有一个形如 $S \to aS$ 的规则，该规则表明字符 S 可被重写为 aS(例如，应用该规则三次会产生字符串 $aaaS$)。如果 a 表示一个圆［见图 11.34(a)］，并且"向右侧添加圆"的含义是分配形如 $aaa\cdots$ 的一个字符串，那么规则 $S \to aS$ 将生成如图 11.34(b)所示的纹理模式。

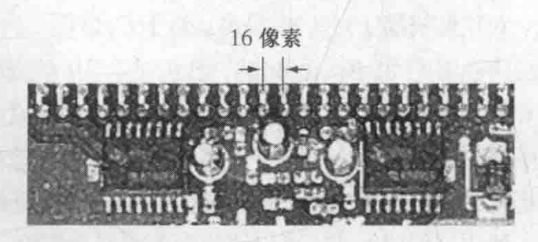


图 11.33　显示有周期分量的电路板局部放大图像

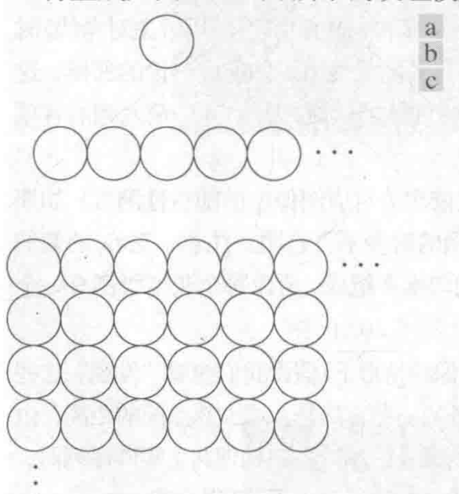


图 11.34　(a)纹理基元；(b)由规则→As 生成的模式；(c)由该规则和其他规则生成的二维纹理模式

假设我们接着向该方案中添加一些新的规则：$S \to bA$，$A \to cA$，$A \to c$，$A \to bS$，$S \to a$，其中 b 表示"下方添加一个圆"，c 表示"向左侧添加一个圆"。我们现在生成一个形如 $aaabccbaa$ 的字符串，它对应于一个圆的 3×3 矩阵。使用相同的方法可以生成更大的纹理模式，如图11.34(c)所示(注意，这些规则也可生成非矩形结构)。

前述讨论的基本思想是，一个简单的"纹理基元"可借助一些规则用于形成更复杂的纹理模式，这些规则限制基元(或这些基元)的可能排列的数量。这些概念是关系描绘的核心，11.5 节中将详细讨论这一主题。

频谱法

如在 5.4 节中讨论的那样，傅里叶频谱非常适合描述图像中的二维周期或近似二维周期模式的方向性。这些全局纹理模式在频谱中作为高能脉冲集中区域很容易区分。这里，我们考虑对纹理描述很有用的傅里叶频谱的三个特征：(1)频谱中突出的尖峰给出纹理模式的主要方向；(2)频率平面中尖峰的位置给出模式的基本空间周期；(3)采用滤波方法消除任何周期成分而留

858 ~ 860

下非周期性图像元素，然后采用统计技术来描述。回忆可知，频谱是关于原点对称的，所以只需要考虑半个频率平面。因此，对于分析的目的，每个周期模式只涉及频谱中的一个尖峰，而不是两个尖峰。

刚才提到的频谱特征的检测和描述通常可以使用极坐标表达来简化：得到一个函数 $S(r, \theta)$，其中 S 是频谱函数，r 和 θ 是该坐标系中的变量。对于每个方向 θ, $S(r, \theta)$ 可以视为一个一维函数 $S_\theta(r)$。类似地，对于每个频率 r，$S_r(\theta)$ 也是一个一维函数。对固定值 θ 分析 $S_\theta(r)$，可得到径向上的频谱特性(譬如出现尖峰)；而对定值 r 分析 $S_r(\theta)$，可得到以原点为圆心的一个圆上的频谱特性。

对这些函数进行积分(对离散变量求和)，可得到一个更为整体的描述：

$$S(r) = \sum_{\theta=0}^{\pi} S_\theta(r) \tag{11.3-10}$$

和

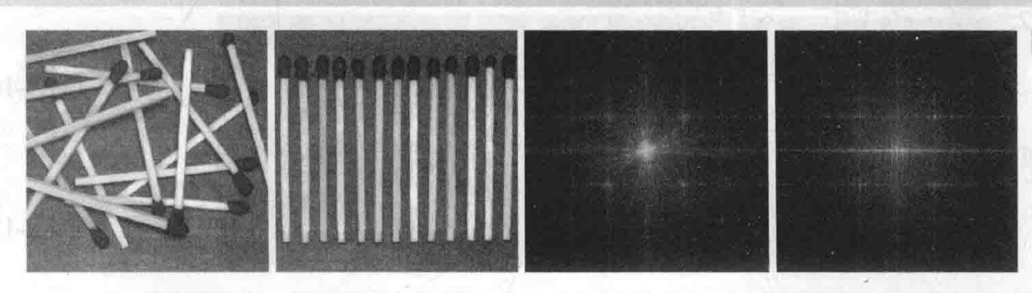
$$S(\theta) = \sum_{r=0}^{R_0} S_r(\theta) \tag{11.3-11}$$

式中，R_0 是以原点为圆心的圆的半径。

式(11.3-10)和式(11.3-11)的结果为每对坐标 (r, θ) 构成了一对值 $[S(r), S(\theta)]$。通过改变这些坐标，我们可以生成两个一维函数 $S(r)$ 和 $S(\theta)$，从而为整幅图像或所考虑区域的纹理构成一种频谱-能量描述。此外，为了定量地表征它们的特性，可以计算出这些函数本身的描绘子。用于这一目的的典型描绘子是最高值的位置、幅度和轴向变化的均值与方差，以及该函数的均值和最高值之间的距离。

例 11.12 频谱纹理。

图 11.35(a)显示了随机分布的火柴图像，图 11.35(b)显示了按周期排列后的火柴图像。图 11.35(c)和(d)显示了相应的傅里叶频谱。两个傅里叶频谱中在二维方向上延伸的四边形周期能量脉冲，归因于火柴的其余部分的粗糙背景的周期纹理。图 11.35(c)中频谱的其他支配成分是由图 11.35(a)中物体边缘的随机方向引起的。另一方面，图 11.35(d)中与背景无关的主要能量是沿水平轴分布的，它对应于图 11.35(b)中的较强垂直边缘。



a b c d

图 11.35 (a)随机分布的火柴图像；(b)有序排列的火柴图像；(c)图(a)的傅里叶频谱；(d)图(b)的傅里叶频谱。所有图像的大小均为 600×600 像素

图 11.36(a)和(b)是随机排列火柴的 $S(r)$ 曲线和 $S(\theta)$ 曲线，类似地，图 11.36(c)和(d)是有序排列火柴的 $S(r)$ 曲线和 $S(\theta)$ 曲线。随机排列火柴的 $S(r)$ 曲线表明没有较强的周期分量[即与原点处的尖峰(直流分量)相比，频谱中没有其他支配尖峰]。相反，有序排列火柴的 $S(r)$ 曲线在 $r = 15$ 附近显示了一个较强的尖峰，而在 $r = 25$ 附近显示了一个较小的尖峰，它们分别对应于图 11.35(b)中亮区域(火柴)和暗区域(背景)的水平重复周期。类似地，图11.35(c)中的能量脉冲的随机特性，在图 11.36(b)所示的 $S(\theta)$ 曲线中十分明显。相比之下，图11.36(d)所示图形在靠近原点、90°和 180°的区域显示了很强的能量分量。这与图11.35(d)中频谱的能量分布是一致的。

860 ~ 861

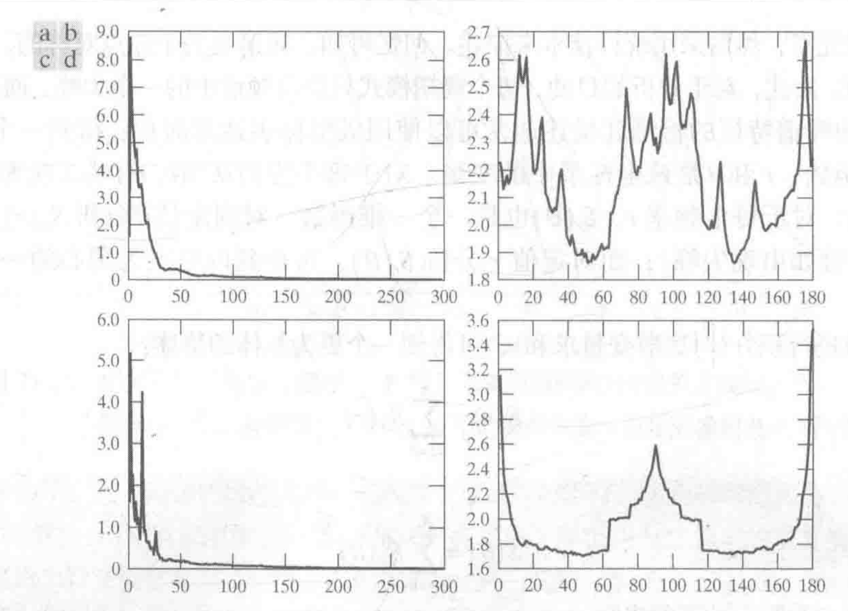


图 11.36 (a)图 11.35(a)的 $S(r)$曲线；(b)图 11.35(a)的 $S(\theta)$曲线；(c)图 11.35(b)的 $S(r)$曲线；(d)图 11.35(b)的 $S(\theta)$曲线。所有垂直轴都要$\times 10^5$

11.3.4 不变矩

大小为 $M\times N$ 的数字图像 $f(x, y)$的二维$(p+q)$阶矩定义为

$$m_{pq}=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}x^p y^q f(x,y) \tag{11.2-12}$$

式中 $p=0,1,2,\cdots$ 和 $q=0,1,2,\cdots$ 是整数。相应的$(p+q)$阶中心矩定义为

$$\mu_{pq}=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}(x-\overline{x})^p(y-\overline{y})^q f(x,y) \tag{11.3-13}$$

式中 $p=0,1,2,\cdots$ 和 $q=0,1,2,\cdots$，其中

$$\overline{x}=\frac{m_{10}}{m_{00}} \quad 和 \quad \overline{y}=\frac{m_{01}}{m_{00}} \tag{11.3-14}$$

由 η_{pq} 表示的归一化中心矩定义为

$$\eta_{pq}=\frac{\mu_{pq}}{\mu_{00}^{\gamma}} \tag{11.3-15}$$

式中，

$$\gamma=\frac{p+q}{2}+1 \tag{11.3-16}$$

其中 $p+q=2,3,\cdots$。

由二阶矩和三阶矩可推出如下 7 个不变矩组[①]：

$$\phi_1=\eta_{20}+\eta_{02} \tag{11.3-17}$$

$$\phi_2=(\eta_{20}-\eta_{02})^2+4\eta_{11}^2 \tag{11.3-18}$$

861 ~ 863

① 推导这些结果时，所用到的概念超出了本书的讨论范围。在由 Bell[1965]撰写的图书和由 Hu[1962]发表的论文中，包含了这些概念的详细讨论。关于生成高于 7 阶不变矩的方法，见 Flusser[2000]。不变矩可推广到 n 维情形(Mamistvalov[1998])。

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{11.3-19}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{11.3-20}$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ & (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{11.3-21}$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{11.3-22}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ & (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{11.3-23}$$

这些矩组对于平移、尺度变化、镜像(内部为负号)和旋转是不变的。

例 11.13　不变矩。

本例的目的是，利用图11.37(a)中的图像计算并比较前述的几个不变矩。黑(0)边界是加上去的，以使本例的所有图像都有相同的尺寸；0 不影响不变矩的计算。图11.37(b)到(f)分别显示了原图像经平移、在空间两个维度上缩小一半、镜像、旋转 45°和旋转 90°操作后的结果。表 11.5 总结了这 6 幅图像的 7 个不变矩的值。为减小动态范围进而简化解释，所显示的值是 $\mathrm{sgn}(\phi_i)\log_{10}(|\phi_i|)$。需要绝对值的原因是，许多值是小数和/或负数；符号函数保留符号(这里我们感兴趣的是矩的不变性和相对符号，而不是它们的真实值)。表 11.5 中的两个关键点是：(1)矩的值的接近程度，与平移、尺度缩放、镜像和旋转无关；(2) ϕ_7 的符号对于镜像图像是不同的(这是实践中用于检测一幅图像是否已被镜像的一种性质)。

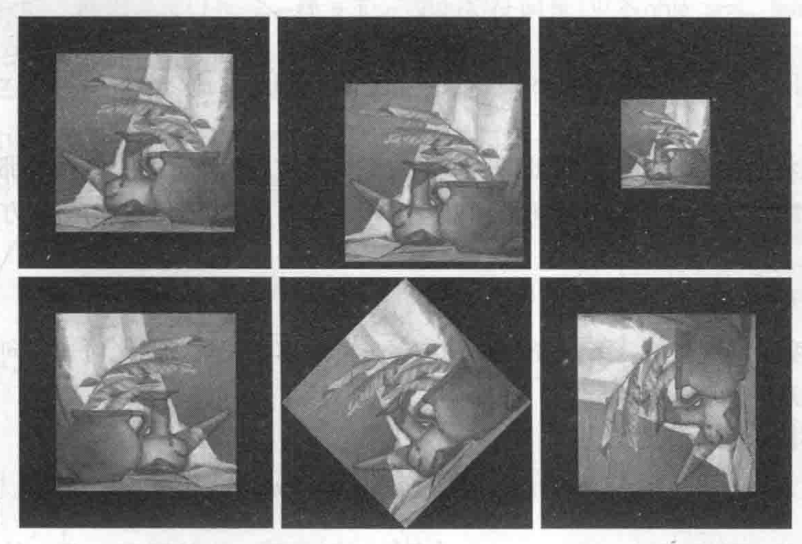


图 11.37　(a)原图像；(b)~(f)分别经平移、缩小一半、镜像、旋转 45°和旋转 90°操作后的图像

表 11.5　图 11.37 中各幅图像的不变矩

不 变 矩	原 图 像	平　移	缩 小 一 半	镜　像	旋转 45°	旋转 90°
ϕ_1	2.8662	2.8662	2.8664	2.8662	2.8661	2.8662
ϕ_2	7.1265	7.1265	7.1257	7.1265	7.1266	7.1265
ϕ_3	10.4109	10.4109	10.4047	10.4109	10.4115	10.4109
ϕ_4	10.3742	10.3742	10.3719	10.3742	10.3742	10.3742
ϕ_5	21.3674	21.3674	21.3924	21.3674	21.3663	21.3674
ϕ_6	13.9417	13.9417	13.9383	13.9417	13.9417	13.9417
ϕ_7	–20.7809	–20.7809	–20.7724	20.7809	–20.7813	–20.7809

863
~
864

11.4 使用主成分进行描绘

本节中讨论的内容适用于边界和区域。另外，这些内容也是描绘一组空间上已配准图像的基础，但这些已配准图像的对应像素值是不同的(即 RGB 图像的三个分量图像)。假设我们已有这样一幅彩色图像的三个分量图像。通过将每组三个对应像素表示为一个向量，可将这三幅图像作为一个单元来处理。例如，令 x_1, x_2 和 x_3 分别是这三幅 RGB 分量图像中的一个像素的值。这三个元素可以表示为一个三维列向量 $\boldsymbol{x}$，即

关于向量和矩阵的简要回顾，请查阅本书的网站。

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

这个向量表示所有三幅图像中的一个共同像素。如果图像大小为 $M\times N$，则按这种方式表示所有像素后，共有 $K=MN$ 个三维向量。如果有 n 幅已配准的图像，那么向量将是 n 维的：

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\\x_3\\\vdots\\x_n\end{bmatrix}\tag{11.4-1}$$

本节假设所有向量都是列向量(即 $n\times1$ 阶矩阵)。我们可将它们写成一行，表示为 $\boldsymbol{x}=(x_1,\ x_2,\ \cdots,\ x_n)^{\mathrm{T}}$ 的形式，其中 T 表示转置。

我们可以将向量当作随机量来处理，就像我们构建灰度级直方图时所做的那样。唯一的差别是，我们现在谈论的是随机向量的均值向量和方差矩阵，而不是随机变量的均值和方差。总体均值向量定义为

$$\boldsymbol{m}_{\mathbf{x}}=E\{\boldsymbol{x}\}\tag{11.4-2}$$

式中 $E\{\cdot\}$是变量的期望值，下标表示 $\boldsymbol{m}$ 与向量 $\boldsymbol{x}$ 的总体相关联。回忆可知，向量或矩阵的期望值可通过取每个元素的期望值来得到。

向量总体协方差矩阵定义为

$$\boldsymbol{C}_{\mathbf{x}}=E\left\{(\boldsymbol{x}-\boldsymbol{m}_x)(\boldsymbol{x}-\boldsymbol{m}_x)^{\mathrm{T}}\right\}\tag{11.4-3}$$

因为 $\boldsymbol{x}$ 是 n 维的，故 $\boldsymbol{C}_x$ 和 $(\boldsymbol{x}-\boldsymbol{m}_x)(\boldsymbol{x}-\boldsymbol{m}_x)^{\mathrm{T}}$ 是 $n\times n$ 阶矩阵。$\boldsymbol{C}_x$ 的元素 c_{ii} 是 x_i 的方差，即总体中 $\boldsymbol{x}$ 向量的第 i 个分量，$\boldsymbol{C}_x$ 的元素 c_{ij} 是这些向量元素 x_i 和 x_j 之间的协方差[①]。矩阵 $\boldsymbol{C}_x$ 是实对称的。如果元素 x_i 和 x_j 是不相关的，则它们的协方差为零，从而有 $c_{ij}=c_{ji}=0$。当 $n=1$ 时，所有这些定义都简化为它们所熟悉的一维形式。

对于来自随机总体的 K 个向量取样，均值向量可以通过使用常见的求平均值的表达式由样本来近似：

$$\boldsymbol{m}_x=\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{m}_x\tag{11.4-4}$$

864 ~ 866

① 回忆可知，均值为 m 的随机变量 x 的方差定义为 $E\{(x-m)^2\}$。两个随机变量 x_i 和 x_j 的协方差定义为 $E\{(x_i-m_i)(x_j-m_j)\}$。

类似地，通过展开乘积$(\boldsymbol{x}-\boldsymbol{m}_x)(\boldsymbol{x}-\boldsymbol{m}_x)^{\mathrm{T}}$并使用式(11.4-2)和式(11.4-4)，我们发现协方差矩阵可由如下样本来近似：

$$\boldsymbol{C}_x=\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}-\boldsymbol{m}_x\boldsymbol{m}_x^{\mathrm{T}} \tag{11.4-5}$$

例 11.14　均值向量和协方差矩阵的计算。

为说明式(11.4-4)和式(11.4-5)的机理，考虑 4 个向量$\boldsymbol{x}_1=(0,0,0)^{\mathrm{T}}$, $\boldsymbol{x}_2=(1,0,0)^{\mathrm{T}}$, $\boldsymbol{x}_3=(1,1,0)^{\mathrm{T}}$和$\boldsymbol{x}_4=(1,0,1)^{\mathrm{T}}$。应用式(11.4-4)得到如下均值向量：

$$\boldsymbol{m}_x=\frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}$$

同样，使用式(11.4-5)得到下列协方差矩阵：

$$\boldsymbol{C}_x=\frac{1}{16}\begin{bmatrix}3&1&1\\1&3&-1\\1&-1&3\end{bmatrix}$$

沿主对角线的所有元素都相等，这表明总体中向量的三个分量有相同的方差。此外，元素x_1和x_2及x_1和x_3是必定相关的，元素x_2和x_3是必定不相关的。

因为$\boldsymbol{C}_x$是实对称的，所以总可以找到一组 n 个正交的特征向量(见 Noble and Daniel[1988])。令$\boldsymbol{e}_i$和λ_i, $i=1,2,\cdots,n$为$\boldsymbol{C}_x$的特征向量和对应的特征值①，为方便起见以降序排列，以使$\lambda_j \geqslant \lambda_j+1, j=1,2,\cdots,n-1$。令$\boldsymbol{A}$为一个矩阵，其行由$\boldsymbol{C}_x$的特征向量组成，特征向量的排序方式为$\boldsymbol{A}$的第一行对应于最大特征值的特征向量，最后一行对应于最小特征值的特征向量。

假设我们将$\boldsymbol{A}$用做将向量$\boldsymbol{x}$映射为向量$\boldsymbol{y}$的一个变换矩阵，如下所示：

$$\boldsymbol{y}=\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{m}_x) \tag{11.4-6}$$

该表达式称为霍特林变换，就像马上要说明的那样，该变换具有一些有趣且有用的性质。

> 霍特林变换与离散 Karhunen-Loève 变换(见 Karhunen[1947])是相同的，因此文献中交替使用这两个名称。

不难证明，由该变换得到的$\boldsymbol{y}$向量的均值为零，即

$$\boldsymbol{m}_y=E\{\boldsymbol{y}\}=\boldsymbol{0} \tag{11.4-7}$$

由基本矩阵理论可知，$\boldsymbol{y}$的协方差矩阵由$\boldsymbol{A}$和$\boldsymbol{C}_x$按如下表达式给出：

$$\boldsymbol{C}_y=\boldsymbol{A}\boldsymbol{C}_x\boldsymbol{A}^{\mathrm{T}} \tag{11.4-8}$$

此外，因为构成$\boldsymbol{A}$的方式的原因，$\boldsymbol{C}_y$是一个对角矩阵，其主对角线上的元素是$\boldsymbol{C}_x$的特征值，即

$$\boldsymbol{C}_y=\begin{bmatrix}\lambda_1 & & & 0\\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n\end{bmatrix} \tag{11.4-9}$$

该协方差矩阵中非对角线上的元素为零，所以向量$\boldsymbol{y}$的元素是不相关的。记住，λ_j是$\boldsymbol{C}_x$的特征值，并

① 根据定义，大小为 $n\times n$ 的矩阵 $\boldsymbol{C}$ 的特征向量和特征值满足关系 $\boldsymbol{C}\boldsymbol{e}_i=\lambda_i\boldsymbol{e}_i, i=1,2,\cdots,n$。

866~867

且对角矩阵中主对角线上的元素是 C_x 的特征值(见 Noble and Daniel[1988])。因此，C_x 和 C_y 具有相同的特征值。

霍特林变换的另一个重要性质是可处理由 y 重建 x 的问题。因为 A 的各行是正交向量，由此得出 $A^{-1} = A^{\mathrm{T}}$，且任何向量 x 都能通过下式由其对应的 y 来恢复：

$$x = A^{\mathrm{T}} y + m_x \tag{11.4-10}$$

然而，假设我们不使用 C_x 的所有特征向量，而由对应于 k 个最大特征值的 k 个特征向量来形成矩阵 A_k，那么将得到一个 $k \times n$ 阶的变换矩阵。然后，向量 y 成为 k 维矩阵，并且式(11.4-10)给出的重建不再是精确(这与我们在 11.2.3 节中使用少量傅里叶系数描述一个边界的过程有些相似)。

使用 A_k 重建的向量为

$$\hat{x} = A_k^{\mathrm{T}} y + m_x \tag{11.4-11}$$

可以证明，x 和 $\hat{x}$ 之间的均方误差由下式给出：

$$\begin{aligned} e_{ms} &= \sum_{j=1}^{n} \lambda_j - \sum_{j=1}^{k} \lambda_j \\ &= \sum_{j=k+1}^{n} \lambda_j \end{aligned} \tag{11.4-12}$$

式(11.4-12)的第一行表明，如果 $k = n$(即如果变换中使用了所有的特征向量)，则误差为零。因为 λ_j 单调递减，式(11.4-12)还表明，通过选取与最大特征值相对应的 k 个特征向量，可使得误差最小。从而，从可以将向量 x 和其近似值 $\hat{x}$ 之间的均方误差降至最小这方面来说，霍特林变换是最佳的。由于使用对应于最大特征值的特征向量这一思想，霍特林变换也称为**主成分变换**。

例 11.15　使用主成分描述图像。

图 11.38 显示了 6 幅多光谱卫星图像，这些图像对应于 6 个谱带：可见蓝光(450~520 nm)、可见绿光(520~600 nm)、可见红光(630~690 nm)、近红外(760~900 nm)、中红外(1550~1750 nm)和热红外(10 400~12 500 nm)。本例的目的是说明如何运用主成分进行图像描述。

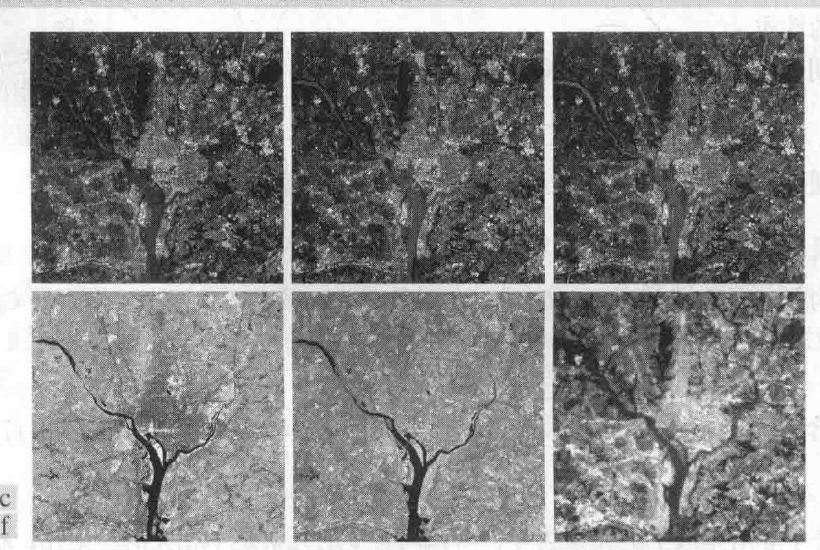


图 11.38　多光谱图像：(a)可见蓝光；(b)可见绿光；(c)可见红光；(d)近红外；(e)中红外；(f)热红外(图像由 NASA 提供)

867~868

如本节开始时讨论的那样，若按图 11.39 中所示的方式来组织图像，那么这些图像中的每组对应像素可形成一个 6 元素的向量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_6)^{\mathrm{T}}$。本例中所有图像的大小都是 564×564 像素，所以一共构成了 $(564)^2$ = 318 096 个向量，由此来计算均值向量、协方差矩阵及相应的特征值和特征向量。然后，特征向量用做矩阵 $\boldsymbol{A}$ 的行，且使用式(11.4-6)得到一组 $\boldsymbol{y}$ 向量。类似地，我们使用式(11.4-8)得到了 $\boldsymbol{C}_y$。表 11.6 显示了该矩阵的特征值。注意前两个特征值的支配性。

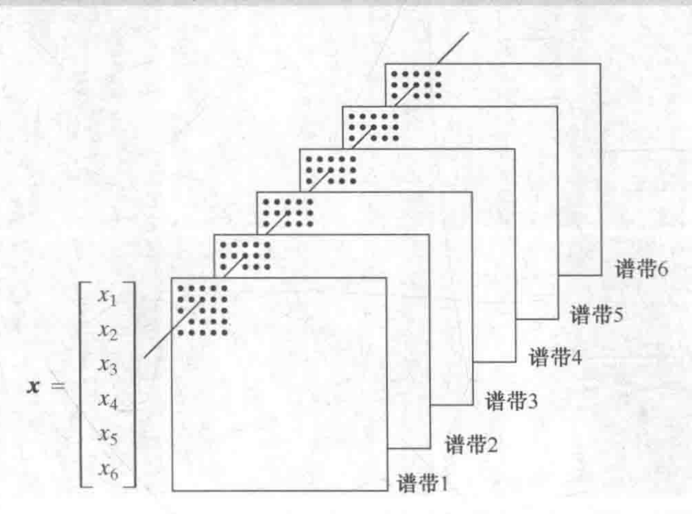


图 11.39　由 6 幅图像中的对应像素形成一个向量

前一段中提到使用 $\boldsymbol{y}$ 向量生成了一组主成分图像(反过来，应用图 11.39 可由这些向量构建图像)。图 11.40 显示了结果。图 11.40(a)是由 318 096 个 $\boldsymbol{y}$ 向量的第一个分量形成的，图 11.40(b)是由这些向量的第二个分量形成的，以此类推。因此，这些图像的大小与图 11.38 中原图像的大小相同。主成分图像中最明显的特征是，对比度细节的重要部分包含在前两幅图像中，并自此快速降低。通过研究特征值，即可说明其原因。如表 11.6 所示，前两个特征值远大于其他特征值。因为特征值是 $\boldsymbol{y}$ 向量的元素的方差，而方差是灰度对比度的测度，不难预料，由对应于最大特征值的向量分量形成的图像会显示最高的对比度。事实上，图 11.40 中的前两幅图像占总方差的 89%。其他 4 幅图像有着较低的对比度细节，因为它们仅占总方差的 11%。

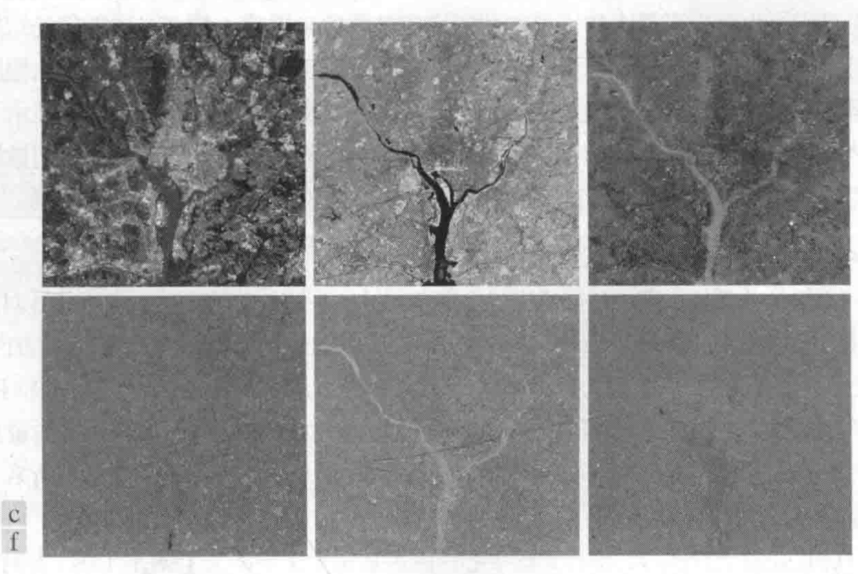


图 11.40　使用式(11.4-6)计算的向量得到的 6 幅主成分图像。反过来，应用图 11.39 可由向量得到图像

869
~
870

表 11.6 由图 11.38 中的各幅图像得到的协方差矩阵的特征值

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
10 344	2966	1401	203	94	31

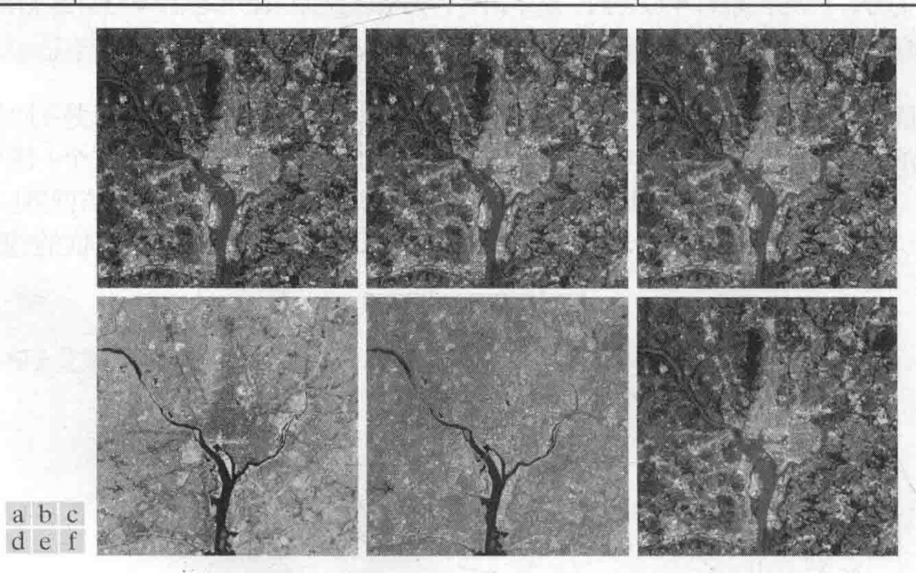


图 11.41 仅使用两幅主成分图像重建的多光谱图像，这些图像对应于有着最大特征值(方差)的两幅主成分图像。请将这些图像与图 11.38 中的原图像进行比较

根据式(11.4-11)和式(11.4-12)，如果使用矩阵 $\boldsymbol{A}$ 中的所有特征向量，那么我们能够在原始图像和重建图像间以 0 误差从主成分图像(向量)重建原始图像(向量)。也就是说，原图像和重建图像是相同的。如果我们的目的是存储和/或传输主成分图像及用于随后重建原图像的变换矩阵，则存储和/或传输所有的主成分图像就没有意义，因为这不会带来任何益处。然而，假设我们仅保留和/或传输两幅主成分图像(它们拥有多数对比度细节)，则在存储和传输上会有较大的节省(矩阵 $\boldsymbol{A}$ 的大小是 2×6，故其影响可忽略不计)。

当我们称呼图像时，会交替使用术语“向量”，因为在此处的上下文中，两者间存在一一对应的关系。

图 11.41 显示了由对应于最大特征值的两幅主成分图像来重建 6 幅多光谱图像的结果。前五幅图像的外观十分接近于图 11.38 中的原图像，但第六幅图像却不是这样。原因是原始的第六幅图像实际上是模糊的，而在重建中使用的主成分图像是清晰的，因此丢失了模糊的“细节”。图 11.42 显示了原图像和重建图像间的差。图 11.42 中的图像已被增强，目的是突出它们之间的差别。如果不对它们做增强处理，那么前五幅图像看起来几乎会是全黑的。如期望的那样，第六幅差值图像显示了最大的不同。

例 11.16 用主成分对尺度、平移和旋转变化归一化。

如本章前面提及的那样，表示和描述应尽可能与尺寸、平移和旋转无关。主成分针对这三个参量的变化提供了归一化边界和/或区域的方便方法。考虑图 11.43 中的物体。假设其尺寸、位置和方向(旋转)是任意的。区域内的(或区域边界上的)点可视为二维向量 $\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}}$，其中 x_1 和 x_2 分别是沿 x_1 轴和 x_2 轴的任意一点的坐标值。区域或边界中的所有点组成一个二维向量总体，如先前那样，这个二维向量总体用于计算协方差矩阵 $\boldsymbol{C_x}$ 和均值 $\boldsymbol{m_x}$。$\boldsymbol{C_x}$ 的第一个特征向量指向总体的最大方差(数据扩散)方向，而第二个特征向量与第一个特征向量垂直，如图 11.43(b)所示。根据当前的讨论，式(11.4-6)中的主成分变换完成两件事情：(1)在总体的质心(均值)处建立变换后的坐标系的中心，因为从每个 $\boldsymbol{x}$ 中减去了 $\boldsymbol{m_x}$；(2)产生的 $\boldsymbol{y}$ 坐标是 $\boldsymbol{x}$ 旋转后的形式，因此数据就与特征向量联系起来。如果我们定义一个 (y_1,y_2) 轴系统，使 y_1 沿第一个特征向量，而

870 ~ 872

y_2 沿第二个特征向量，那么导致的几何形状如图11.43(c)所示。也就是说，占支配地位的数据的方向与轴系统一致。不管物体的尺度缩放、平移或旋转，只要区域或边界中的所有点都经历相同的变化，就会得到相同的结果。如果我们希望在尺度上归一化变换后的数据，那么可使用对应的特征值来除该坐标。

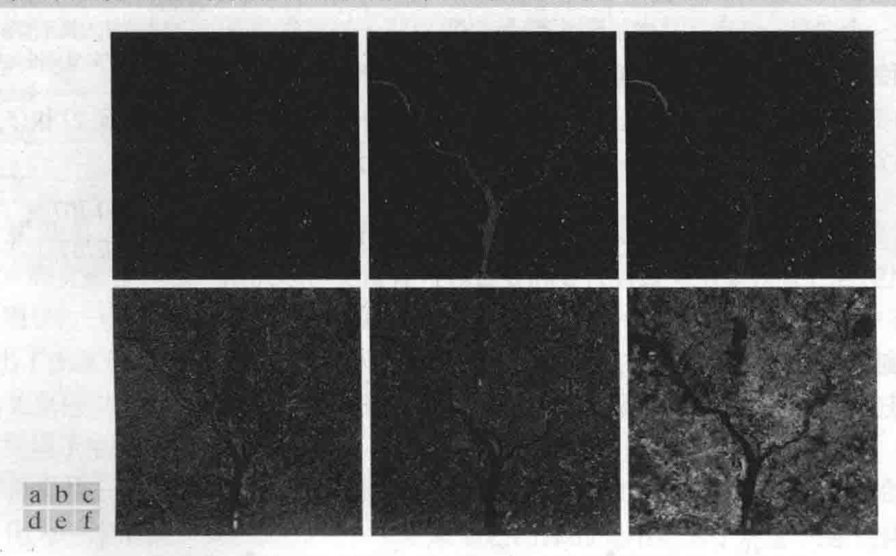


图 11.42　原图像与重建图像之间的差。为便于视觉分析，所有差值图像都已被增强，即已将它们标定到整个范围[0, 255]内

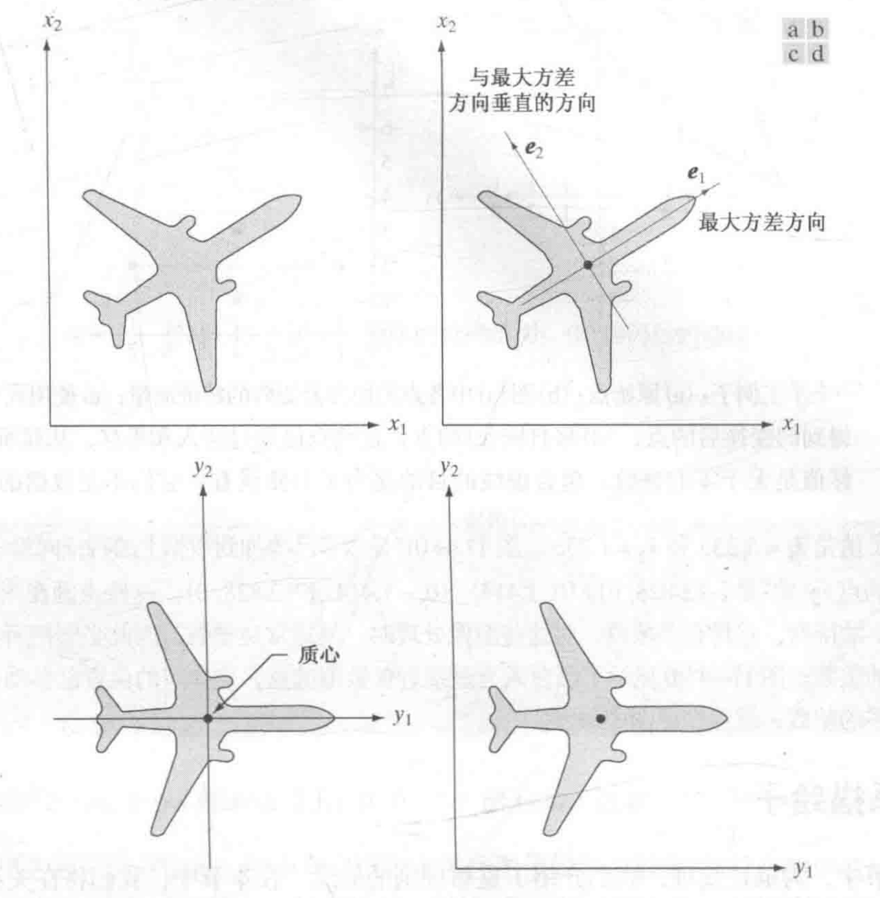


图 11.43　(a)一个物体；(b)显示其协方差矩阵的特征向量的物体；(c)使用式(11.4-6)得到的变换后的物体；(d)平移后的物体，其所有的坐标值都大于零

872
~
873

观察图 11.43(c)可知，y 轴系统中的点可能既有正值又有负值。为将所有的坐标转换为正值，我们简单地从所有 $\mathbf{y}$ 向量中减去 $(y_{1\min}, y_{2\min})^{\mathrm{T}}$。为置换结果点以便使它们都大于零，就像在图11.43(d)中那样，我们对所有的点都加上一个向量 $(a,b)^{\mathrm{T}}$，其中 a 和 b 都大于 0。

> y 轴系统的方向可以是与图 11.43(c)所示方向相反的 180°，具体取决于原始物体的方向。例如，如果图 11.43(a)中的机鼻指向相反的方向，则导致的特征向量会指向左下方。

尽管前面的讨论原理上很简单，但其机理常常会引起混淆。故我们使用一个简单的手工示例来结束该例子。图 11.44(a)显示了坐标分别为(1, 1)，(2, 4)，(4, 2)和(5, 5)的 4 个点。均值向量、协方差矩阵和该总体的归一化(单位长度)特征向量是

$$\boldsymbol{m}_x = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \qquad \boldsymbol{C}_x = \begin{bmatrix} 3.333 & 2.00 \\ 2.00 & 3.333 \end{bmatrix}, \qquad \boldsymbol{e}_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}, \qquad \boldsymbol{e}_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$$

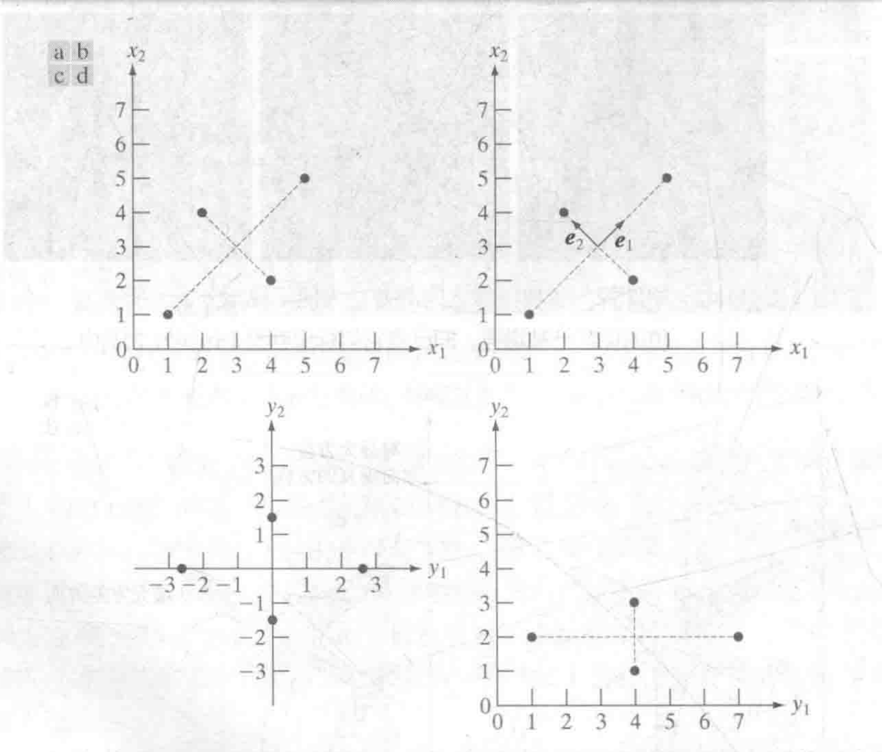


图 11.44 一个手工例子：(a)原始点；(b)图(a)中各点的协方差矩阵的特征向量；(c)使用式(11.4-6)得到的变换后的点；(d)来自图(c)的点，这些点已经过舍入和平移，从而所有的坐标值是大于零的整数。包含虚线的目的是为了方便观看，它们不是数据的一部分

对应的特征值是 $\lambda_1 = 5.333$ 和 $\lambda_2 = 1.333$。图 11.44(b)显示了已叠加到数据上的特征向量。由式(11.4-6)可知，变换后的点($\mathbf{y}$ 项)是(−2.828, 0)，(0, 1.414)，(0, −1.414)和(2.828, 0)。这些点画在图 11.44(c)中。注意，它们沿 y 轴排列，并具有小数值。对这些图像处理时，值通常是整数，因此必须把所有的分数四舍五入为最接近的整数。图 11.44(d)显示了已舍入为最接近整数值的点，这些点的位置已移动，因此所有的坐标值是大于零的整数，就像在原图中那样。

11.5 关系描绘子

在 11.3.3 节中，为描述纹理，我们介绍了重写规则的概念。在本节中，我们将在关系描绘子的角度上扩展这一概念。这些内容对边界或区域同样适用，并且它们的主要目的是以重写规则的形式来获取边界或区域中的基本重复模式。

873 ~ 874

考虑图 11.45(a)中所示的简单阶梯结构。假设该结构已从一幅图像中分割出来，并且我们希望以某种形式化方法来描述它。通过定义如图所示的两个基元 a 和 b，我们可以按图 11.45(b)所示的形式对图 11.45(a)进行编码。编码后结构的最明显性质是元素 a 和 b 的重复。所以，一种简单的描述方法是用公式表示涉及这些基元的递归关系。一种可能性是使用重写规则：

(1) $S \to aA$

(2) $A \to bS$

(3) $A \to b$

其中 S 和 A 是变量，元素 a 和 b 是对应于刚才定义的基元的常量。规则(1)表明，称为起始符号的 S 可以被基元 a 和变量 A 代替。依次地，变量 A 可以被 b 和 S 代替或只被 b 代替。若使用 bS 来代替 A，便会回到规则(1)，从而重复这一过程。用 b 代替 A 将结束这一过程，因为表达式中不再存在变量。图 11.46 给出了由这些规则导出的一些例子，其中结构下面的数字表示规则 1, 2 和 3 被应用的顺序。a 和 b 之间的关系得以保留，因为这些规则迫使一个 a 后总是跟着一个 b。值得注意的是，这三个简单的重写规则可用于生成(或描述)无限多的“相似”结构。

因为字符串是一维结构，所以将字符串应用于图像描述时，需要建立一种适当的方法来将二维位置关系简化为一维形式。大多数使用字符串来描述图像的应用均基于从感兴趣物体中提取连接线段的思想。一种方法是追踪一个物体的轮廓，并使用指定方向和/或长度的线段来对结果进行编码。图 11.47 说明了这一过程。

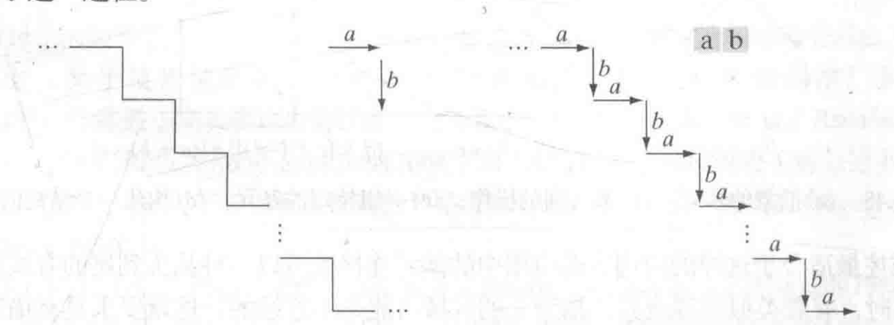


图 11.45　(a)一个简单的阶梯结构；(b)编码后的结构

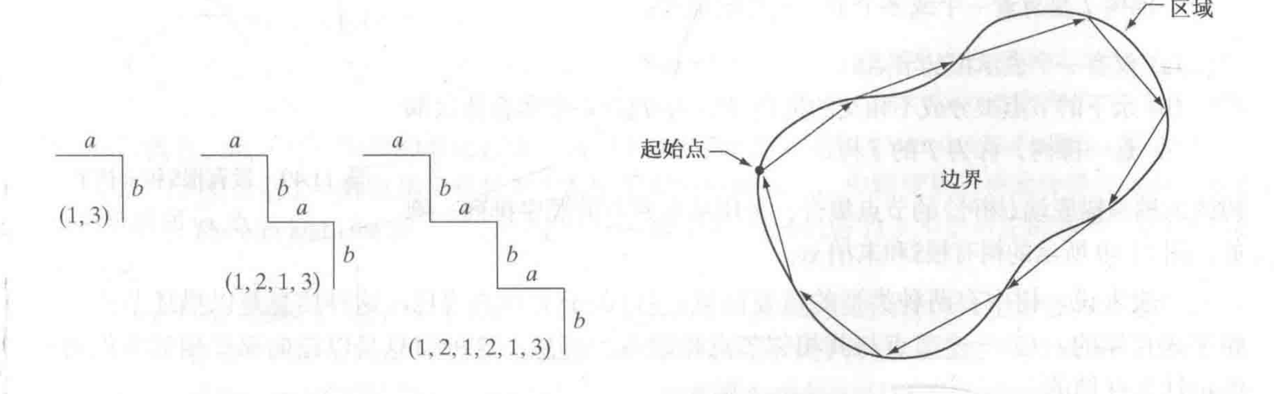


图 11.46　由规则 $S \to aA, A \to bS$ 和 $A \to b$ 导出的例子

图 11.47　使用有向线段对一个区域边界进行编码

另一种更为通用的方法是，使用有向线段来描述图像的各个部分(如较小的单色区域)，除了采用首尾相连的方式外，这种有向线段也可采用其他的方式来连接。图11.48(a)说明了这一方法，图11.48(b)显示了一些可根据抽象的基元来定义的典型操作。图 11.48(c)显示了一组由四个方向定义的线段组成

874~877

的特定基元。图11.48(d)显示了一个特定形状的生成步骤，其中(~d)表示与基元d方向相反的基元。注意，每个合成的结构只有一个头和一个尾。有意义的结果是最后一个字符串，它描述了完整的结构。

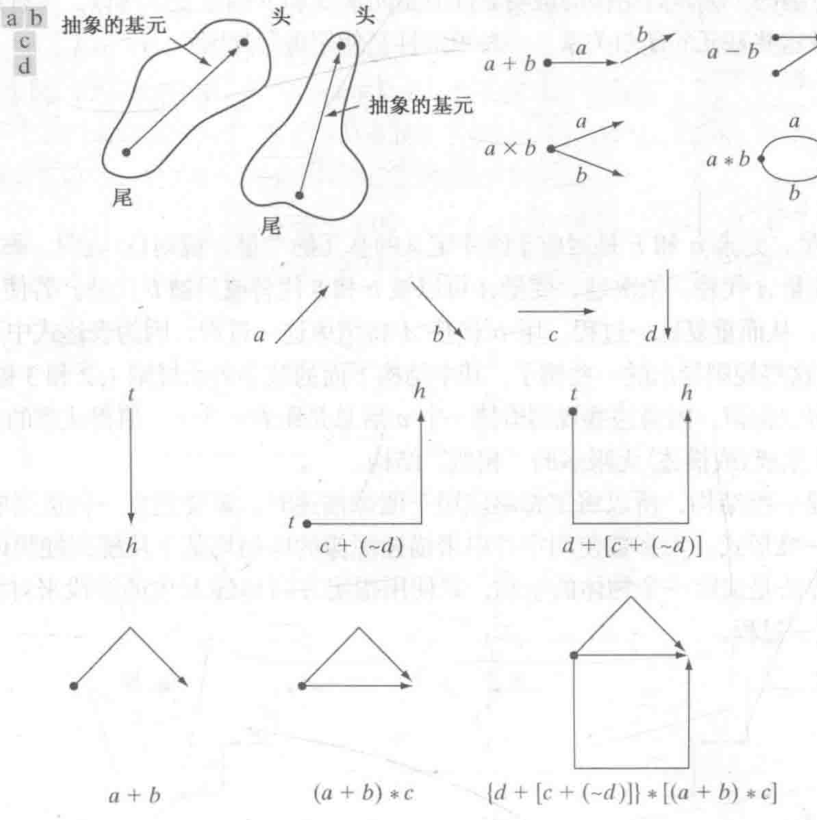


图 11.48　(a)抽象的基元；(b)基元间的操作；(c)一组特定的基元；(d)构建一个结构的步骤

字符串描述最适合于这样的应用，即应用中的基元连接性可以一种从头到尾的方式或其他连续方式来表示。有时，有着类似纹理或其他描绘子的区域可能是不连续的，这就要求具有描述这种情形的技术。处理这种情况的有用方法之一是使用树描绘子。

一棵树T是有着一个或多个节点的有限集合：

(a) 仅有一个表示根的节点\$。

(b) 余下的节点被分成不相交的集合 $T_1,\cdots, T_m$，每个集合依次都是一棵树，称为T的子树。

\$　z　y　x

图 11.49　带有根\$和末梢节点 xy 的简单树

树的末梢是树底端(树叶)的节点集合，采用从左到右的顺序排列。例如，图 11.49 所示的树有根\$和末梢xy。

一般来说，树中有两种类型的重要信息：(1)关于节点的信息，这种信息是以描述节点的一组字来存储的，(2)一个节点与其相邻节点相联系的信息，这种信息是以指向那些相邻节点的一组指针来存储的。

用于图像描述时，第一类信息识别一幅图像子结构(如区域或边界线段)，而第二类信息定义这个子结构与其他子结构间的物理关系。例如，图 11.50(a)可以表示为一棵树，方法是使用关系“在……之内”。因此，如果树的根用\$代表，那么图 11.50(a)表明，第一级复杂性涉及\$内的a和b，这将产生从根发出的两个分支，如图 11.50(b)所示。下一级涉及a内的b以及c内的d和e。最后，e内的f完成整棵树。

877

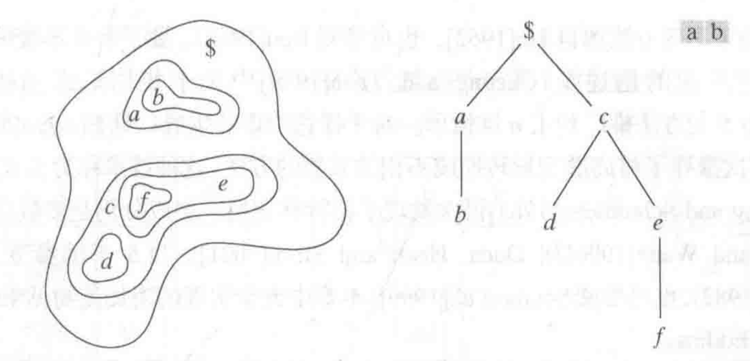


图 11.50　(a)一个简单的合成区域；(b)使用关系“在……之内”得到的树表示

小结

从图像中分割出物体或区域的表示和描绘，是大多数图像自动处理操作的前期步骤。例如，这些描述构成了下一章讨论的目标识别方法的输入。正如本章涵盖的描绘技术指出的那样，具体选择其中的哪一种方法，将取决于所考虑的问题。目的是选择能够“捕获”物体或物体类之间的本质差异的描绘子，同时尽可能保留位置、大小和方向的独立性。

参考文献

11.1.1 节讨论的边界追踪算法由 Moore[1968]首次提出。11.1.2 节讨论的链码表示由 Freeman[1961, 1974]首次提出。关于使用链码的当前研究见 Bribiesca[1999]，他还将链码推广到了三维情形(Bribiesca[2000])。计算最小周长多边形的详细讨论和算法(11.1.3 节)见 Klette and Rsenfeld[2004]，还可参见 Sloboda et al.[1998]和 Coeurjolly and Klette[2004]。11.1.4 节内容的其他重要主题包括不变多边形拟合(Voss and Suesse [1997])、多边形近似算法性能的评估方法(Rosin[1997])、一般实现(Huang and Sun[1999])和计算速度(Davis [1999])。

关于标记图(见 11.1.5 节)的参考文献是 Ballard and Brown[1982]和 Gupta and Srinath[1988]。寻找凸壳和凸缺(见 11.1.6 节)的基本公式见 Preparata and Shamos[1985]，也可参阅论文 Liu-Yu and Antipolis[1993]。Katzir et al.[1994]讨论了部分闭合曲线的检测。Zimmer et al.[1997]讨论了计算凸壳的一种改进算法，Latecki and Lakämper[1999]讨论了形状分解的凸性规则。

11.1.7 节中讨论的骨架化算法源于 Zhang and Suen[1984]。关于该算法的性质和实现的另外一些有用评述，见 Lu and Wang[1986]。论文 Jang and Chin[1990]关联了 11.1.7 节的讨论和 9.5.5 节中的细化形态学概念。对于存在噪声的细化方法，见 Shi and Wong[1994]和 Chen and Yu[1996]。Shaked and Bruckstein[1998]讨论了一种可用于从骨架中消除毛刺的修剪算法。中轴变换的快速计算见 Sahni and Jenq[1992]和 Ferreira and Ubéda[1999]。综述性论文 Loncaric[1998]是 11.1 节中讨论的多种方法的重要参考文献。

Freeman and Shapira[1975]给出了寻找一条闭合链码曲线的基本矩形的算法(见 11.2.1 节)。11.2.2 节中关于形状数的讨论基于 Bribiesca and Guzman[1980]和 Bribiesca[1981]。关于傅里叶描绘子(见 11.2.3 节)的其他读物，见 Zahn and Roskies[1972]和 Persoon and fu[1977]，也可参阅 Aguado et al.[1998]和 Sonka et al.[1999]。Reddy and Chatterji[1996]讨论了使用 FFT 算法实现平移不变、旋转不变和缩放不变的一种重要方法。11.2.4 节中的内容是以基本概率理论为基础的(例如，见 Peebles[1993]和 Popoulis[1991])。

关于 11.3.2 节的其他读物，见 Rosenfeld and Kak[1982]和 Ballard and Brown[1982]。关于纹理(见 11.3.3 节)优秀介绍，见 Haralick and Shapiro[1992]。早期关于纹理的综述，见 Wechsler[1980]。Murino et al.[1998]、Garcia[1999]、Shapiro and Stockman[2001]是该领域当前有代表性的成果。

878

11.3.4 节中讨论的矩不变方法源自 Hu[1962]，也可参阅 Bell[1965]。要了解矩不变的应用范围，可参阅 Hall[1979]中关于图像匹配的论述及 Cheung and Teoh[1999]中关于使用矩来描述对称性的论述。Mamistvalov[1998]将矩不变方法推广到了 n 维情形。对于任意阶矩的生成，见 Flusser[2000]。

Hotelling[1933]首次推导了将离散变量转换成不相关系数的方法。这种技术称为*主成分方法*。Kramer and Mathews[1956]和 Huang and Schultheiss[1963]再次发现了霍特林变换。主成分仍是多数应用中描述图像的基本工具，详见 Swets and Weng[1996]和 Duda, Heart and Stork[2001]。11.5 节的参考文献来自 Gonzalez, Thomason[1978]和 Fu[1982]，也可参阅 Sonda et al.[1999]。本章中关于实现的附加读物见 Nixon and Aguado[2002]和 Gonzlez,Woods and Eddins。

习题

11.1 ★(a) 重新定义链码的一个起始点，以便所得的数字序列形成一个最小值整数。请证明该编码与边界上的初始起点无关。

(b) 求编码 10176722335422 的归一化起始点。

11.2 (a) 如 11.1.2 节中解释的那样，证明链码的一次差分会将该链码关于旋转归一化。

(b) 计算编码 0110233210332322111 的一次差分。

11.3 ★(a) 证明 11.1.3 节中讨论的橡皮模多边形近似方法会生成一个最小周长多边形。

(b) 证明：若每个单元对应边界的一个像素，则该单元中的最大可能误差为 $d/\sqrt{2}$，其中 d 是相邻像素之间的最小水平距离或垂直距离(用于生成数字图像的取样网格中线间的距离)。

11.4 在如下条件下，解释 11.1.3 节中的 MPP 算法是如何工作的：

★(a) 1 像素宽、1 像素深的锯齿。

★(b) 1 像素宽、2 或多像素深的锯齿。

(c) 1 像素宽、1 像素长的凸起。

(d) 1 像素宽、n 像素长的凸起。

11.5 ★(a) 如果在 11.1.4 节讨论的聚合方法中将误差阈值设为 0，请讨论对所得多边形的影响。

(b) 如果在分裂方法中这样做，有何影响?

11.6 ★(a) 使用 11.1.5 节讨论的正切角方法，画出一个方形边界的标记图。

(b) 使用斜率密度函数，重复上一问题。

假设该方形分别与 x 轴和 y 轴对准，且 x 轴为参考线。从最靠近原点的角开始。

11.7 求下列每一边界的标记图的表达式，并画出标记图。

★(a) 一个等边三角形。

(b) 一个正方形。

(c) 一个椭圆。

11.8 画出下列图形的中轴：

★(a) 一个圆。

★(b) 一个等边五边形。

(c) 一个矩形。

(d) 一个等边三角形。

879 ~ 880

11.9 对于如下所示的每个图形：

★(a) 讨论在点 p 执行 11.1.7 节中介绍的骨架化算法的步骤 1 时，所产生的作用。

(b) 重复该算法的步骤 2。假设所有情形下 $p=0$。

1	1	0
1	p	0
1	1	0

0	0	0
1	p	0
0	0	0

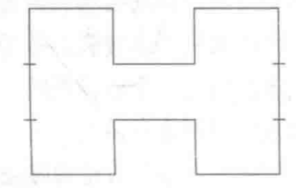
0	1	0
1	p	1
0	1	0

1	1	0
0	p	1
0	0	0

11.10 参考 11.1.7 节中介绍的骨架化算法。请问在执行如下操作后所示图形是什么样子？

★(a) 执行完一遍该算法的步骤 1。

(b) 执行完一遍该算法的步骤 2（在步骤 1 的结果上而非原图像上执行）。



11.11 ★(a) 下图中的形状数的顺序是什么？

(b) 求形状数。

11.12 11.2.3 节中讨论的使用傅里叶描绘子的过程如下：将一个轮廓的坐标表示为复数，求这些复数的 DFT，并只保留 DFT 的一些成分来作为该边界形状的描绘子。求取反 DFT 即可得到原始轮廓的近似。请问哪类轮廓形状具有由实数构成的 DFT？必须怎样设置图 11.19 中的轴系统，才能得到这些实数？

11.13 证明：若在式(11.2-5)中仅用两个傅里叶描绘子($u=0$ 和 $u=1$)来重建一条边界，则结果总是一个圆。（提示：在复平面中使用圆的参数表示，并在极坐标中表示圆的方程。）

★**11.14** 给出区分图 11.10 中所示图形符号所需的统计矩描绘子的最小数量。

11.15 给出具有相同均值和三阶统计矩描绘子但具有不同二阶矩的两个边界形状。

★**11.16** 提出一组能区分字符 0, 1, 8, 6 和 Z 的形状的描绘子。（提示：联合使用拓扑描绘子和凸壳）。

11.17 考虑一幅大小为 100×100 像素的二值图像，图像中有一条从第 1 列延伸到第 99 列的垂直黑色条带，及从第 50 列延伸到第 100 列的垂直白色条带。

(a) 使用位置算子“右侧一个像素”求该图像的共生矩阵。

★(b) 如在 11.3.1 节中解释的那样，归一化该矩阵，以使其元素成为概率估计。

(c) 使用来自(b)的矩阵计算表 11.3 中的 6 个描绘子。

11.18 考虑一幅其中交替出现黑色和白色方块的棋盘图像，每个方块的大小为 $m\times m$。给出可以生成一个对角共生矩阵的位置算子。

11.19 在如下条件下，求一幅其中交替出现 1 和 0 且大小为 8×8 的棋盘图像的灰度级共生矩阵：

★(a) 位置算子 Q 定义为“右侧一个像素”。

(b) 位置算子 Q 定义为“右侧两个像素”。

假设左上角像素的值为 0。

11.20 证明式(11.4-7)、式(11.4-8)和式(11.4-9)的正确性。

★**11.21** 例 11.13 中曾提到，仅用与最大特征值相联系的两个主成分图像，就可重建 6 幅原图像的近似是较为可靠的做法。这样做对均方误差有何影响？请将你的答案表示为最大可能误差的百分比。

880 ~ 881

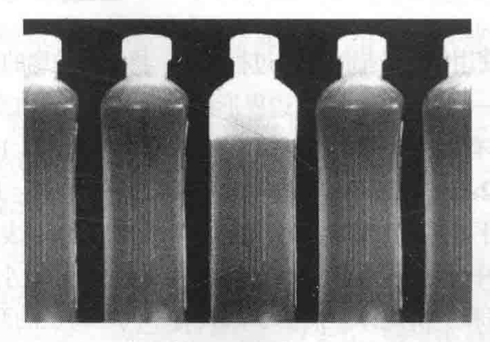
11.22 对于一组大小为 32×32 的图像，假设式(11.4-9)给出的协方差矩阵证明是单位矩阵。原图像和仅使用一半的原始特征向量由式(11.4-11)重建的图像间的均方误差将是什么？

★**11.23** 请问在什么条件下，你会期望 11.2.1 节中定义的边界主轴等于该边界的本征轴？

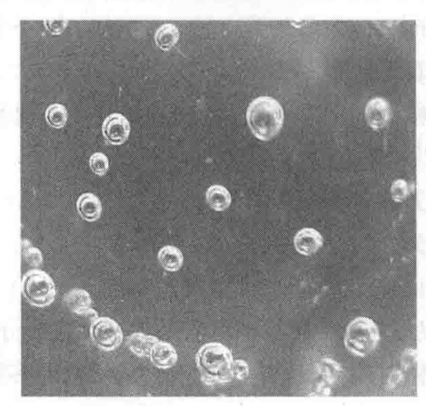
11.24 给出黑白方块棋盘图案的空间关系和对应的树形表达。假设左上角元素为黑色，并且树根对应于该元素。要求从所构造树的每个节点引出的分支不超过两个。

★**11.25** 假设一家公司与你签订了设计一个图像处理系统的合同。图像处理系统用于检测某些实心塑料圆片内部的缺陷。圆片使用一台 X 射线成像系统检测，所生成的 8 比特图像的大小为 1024×1024。不存在缺陷时，图像呈现的是平均灰度为 100、方差为 200 的均匀图像。存在缺陷时，缺陷在图像中表现为斑点状区域，这些区域中约有 70%的像素的灰度偏离 40 这一灰度级，或者均值小于 100。如果这样一个区域的面积超过了 20×20 像素，则认为此圆片有缺陷。请提出一个基于纹理分析的系统。

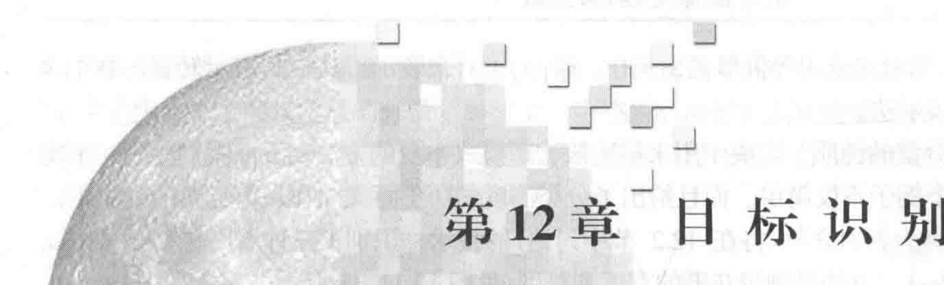
11.26 一家使用瓶子盛装各种工业化学品的公司，在听说你成功地解决了图像处理问题后，雇用你来设计一种检测瓶子未装满的方法。瓶子在传送带上移动并通过自动装填和封盖机时的情形如下图所示。当液位低于瓶颈底部和瓶子肩部的中间点时，则认为瓶子未装满。瓶子横断面的侧面与倾斜面的区域定义为瓶子的肩部。瓶子在不断移动，但该公司有一个成像系统，该系统装备了一个前端照明闪光灯，可有效地停止瓶子的移动，所以你可以得到非常接近于这里显示的样例图像。基于上述资料，请你提出一个检测未完全装满的瓶子的解决方案。清楚地陈述你所做的那些可能会影响到解决方案的所有假设。



11.27 听说你成功地解决了装瓶问题后，一家流体公司与你联系，希望能以某些处理解决气泡自动计数工作，以便进行质量控制。该公司已解决了成像问题，并可以得到分辨率为 800×800 像素的 8 比特图像，如下图所示。每幅图像表示大小为 8 cm^2 的面积。公司希望使用每幅图像来做两件事情：(1)确定图像中气泡所占面积与图像总面积的比率，(2)计算不同气泡的数量。基于上述资料，请你提出该问题的一个解决方案。在你的解决方案中，必须指出可检测到的最小气泡的物理尺寸。清楚地陈述你所做的那些可能会影响到解决方案的所有假设。



881 ~ 882



第12章 目标识别

One of the most interesting aspects of the world is that it can be considered to be made up of patterns.
A pattern is essentially an arrangement. It is characterized by the order of the elements of which it is made, rather than by the intrinsic nature of these elements.

Norbert Wiener

引言

我们通过介绍几种目标识别技术，来结束对数字图像处理的全部讨论。正如 1.1 节中所解释的那样，在界定数字图像处理的覆盖范围时，包含了图像中各个区域的识别，本章中称这些区域为目标或模式。

本章中介绍的模式识别方法主要分为两大领域：决策理论方法和结构方法。第一类方法处理的是使用定量描绘子来描述的各种模式，如长度、面积和纹理等。第二类方法处理的是由定性描绘子来描述的各种模式，如 11.5 节中讨论的关系描绘子。

识别的核心问题是通过样本模式进行“学习”这一概念。下面我们将对决策理论方法和结构方法的学习技术进行深入讨论。

12.1 模式和模式类

模式是描绘子的组合，如第 11 章中讨论的那些描绘子。在有关模式识别文献中经常使用特征来表示描绘子。模式类是指具有某些共同属性的一族模式。模式类用$\omega_1, \omega_2, \cdots, \omega_W$表示，其中 W 是模式类数。由机器完成的模式识别是对不同的模式赋予不同类别的技术，这种技术是自动的，并且能尽可能地减少人为干预。

实践中常用的三种模式组合是向量(用于定量描述)、串和树(用于结构描述)。模式向量由粗体小写字母表示，如 $\boldsymbol{x}, \boldsymbol{y}$ 和 $\boldsymbol{z}$，并采取下列形式：

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{12.1-1}$$

式中，每个分量 x_i 表示第 i 个描绘子，n 是与该模式有关的描绘子的总数。模式向量以列向量(即 $n\times1$ 阶矩阵)的形式表示。因此，模式向量

关于向量和矩阵的简要回顾，请查阅本书的网站。

883
~
884

可以表示为式(12.1-1)所示的形式或用等价形式 $\boldsymbol{x}=(x_1, x_2,\cdots, x_n)^{\mathrm{T}}$ 来表示，其中 T 表示转置。在 11.4 节中，读者应已认识这种表示法。

模式向量 $\boldsymbol{x}$ 中的各个分量的性质，取决于用来描述该物理模式本身的方法。下面我们使用一个简单的例子来加以说明。这个例子不仅简单，而且给出了分类测度的历史意义。在一篇经典的论文中，Fisher[1936]使用一种后来称为判别分析(将在 12.2 节中讨论)的技术，识别了三种鸢尾花(*Iris setosa*, *Iris virginica* 和 *Iris versicolor*)，方法是测量花瓣的宽度和长度(见图 12.1)。

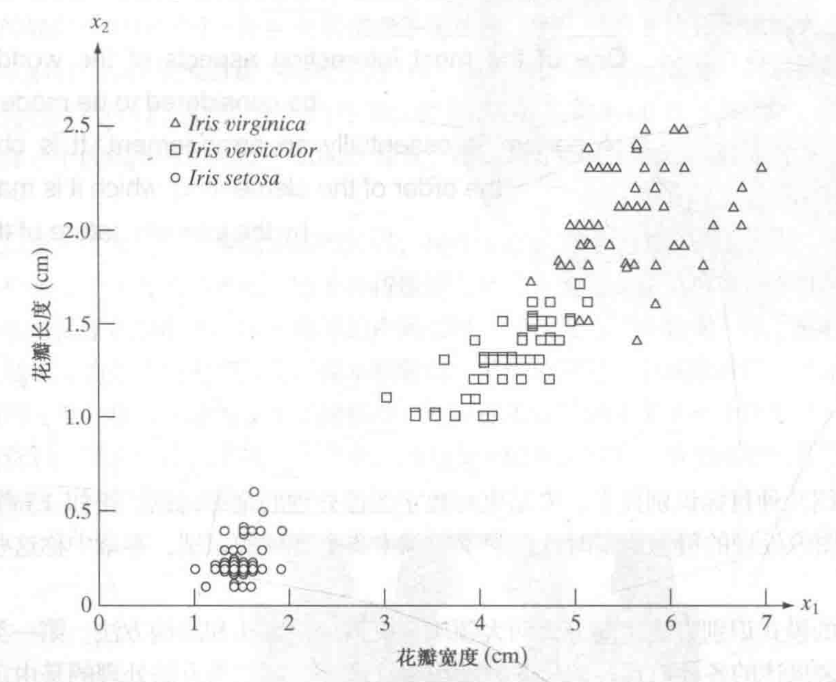


图 12.1 由两个测度描述的三种鸢尾花

在当前术语中，每种花由两个测度来描述，从而生成了形如

$$\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{12.1-2}$$

的一个二维模式，其中 x_1 和 x_2 分别代表花瓣的长度和宽度。在这种情形下，表示为ω_1, ω_2 和ω_3 的 3 个模式类分别对应于 *setosa*, *virginica* 和 *versicolor* 三种花。

由于花瓣在宽度和长度上的不同，描述这些花的模式向量也会不同，这种不同不仅体现在不同的类之间，也体现在一个类的内部。图 12.1 显示了每种鸢尾属花的几个样本的长度和宽度测度。在选定了一组测度后(本例子中为两个测度)，模式向量的分量便成为每个物理样本的完整描述。因此，这种情形下的每朵花就成为二维欧氏空间中的一个点。我们还注意到，本例中花瓣的长度和宽度较好地分离了 *Iris setosa* 与其他两种花，但 *Iris virginica* 和 *Iris versicolor* 的分离并不成功。这一结果说明了经典的特征选择问题，即类的可分程度很大程度上取决于所用描绘子的选择。我们将在 12.2 节和 12.3 节详细阐述这一问题。

图 12.2 显示了模式向量生成的另一个例子。在该例中，我们关注的是不同类型的噪声波形，其中一个样本示于图 12.2(a)。如果选择用其信号(见 11.1.5 节)来表示每个目标，则会得到图 12.2(b)所示形式的一维信号。假设我们用其取样后的幅度值来描述每个信号；也就是说，我们以指定的间隔值 θ 对信号进行取样(表示为 $\theta_1, \theta_2, \cdots, \theta_n$)。然后，我们通过令 $x_1=r(\theta_1), x_2=r(\theta_2), \cdots, x_n=r(\theta_n)$，可构成模式向量。这些向量成为 n 维欧氏空间中的点，并且模式类可以想象为 n 维空间中的“云团”。

884 ~ 885

替代直接使用信号的幅度，我们可以计算一个给定信号的前 n 个统计矩（见 11.2.4 节），并使用这些描绘子作为每个模式向量的分量。事实上，生成模式向量的其他方法还有许多种。本章中我们将介绍现有的一些方法。目前，应记住的关键概念是模式向量中每个分量的描绘子的选择，描绘子对基于模式向量方法的目标识别的最终性能影响巨大。

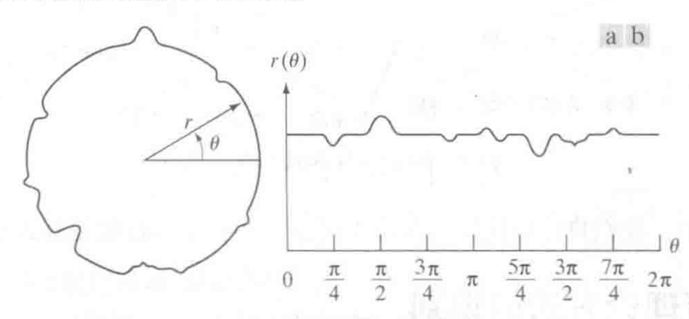


图 12.2　一个带有噪声的目标及其相应的信号

刚才描述的用于生成模式向量的技术，可得到由定量信息表征的模式类。在某些应用中，模式特性可由结构关系来很好地描述。例如，指纹识别基于痕迹特性的相互关系，这种相互关系称为细节。与它们的相对尺寸和位置一起，这些特性是描述指纹脊线特性的主要分量，如指纹的断点、分支、合并和不连续线段。这类识别问题使用结构方法通常能得到解决，因为这种问题不仅定量度量每种特性，而且这些特性间的空间关系决定了类别的成员。11.5 节中已介绍过这一主题。这里，我们从模式描绘子的意义上，再次对其进行简单介绍。

图 12.3(a) 显示了一个简单的阶梯模式。采用类似于图 12.2 中使用的方法，该模式可被取样并表示为一个模式向量。然而，如果采用这种描述方法，那么由两个简单的主要元素重复组成的基本结构将会丢失。一种更有意义的描述方法是定义元素 a 和 b，并将该模式定义为如图 12.3(b) 所示的串 $w = \cdots abababab \cdots$。在这种描述中，该特殊模式类的结构是按如下方式得到的：以首尾相连的方式定义连接性，且只允许符号交替。这种结构组成适用于任何长度的阶梯，但排除了由基元 a 和 b 的其他组合生成其他类型的结果。

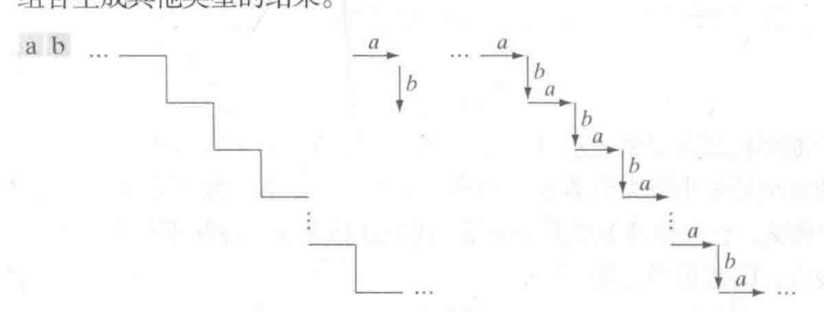


图 12.3　(a) 阶梯结构；(b) 使用基元 a 和 b 对结构编码，生成串描述 $\cdots ababab \cdots$

图 12.4　建筑物密集的城市中心区（华盛顿特区）和周围居民区的卫星图像（原图像由 NASA 提供）

串描述适于生成其结构是基于基元的简单连接，并且通常是和边界形状有关系的目标模式和其他实体模式。对许多应用来说，一种更有效的方法是使用树形描述，如 11.5 节中定义的那样。基本上，多数层次排序方案都会导致树结构。例如，图 12.4 显示了建筑物密集的市区和周围居民区的一幅卫星图像。我们使用符号$来定义整个图像区域。图 12.5 中（从上到下）显示的树形表示是通过使用结构关系“由……组成”得到的。因此，树的根表示整幅图像。下一层表示该图像由市区和居民区构成。居民区依次由住宅、高速公路和购物中心构成。再下一层描述住宅和高速公路。我们可以继续按这样的方式细分，直到达到我们在图像上解析不同区域的能力的极限为止。

885 ~ 887

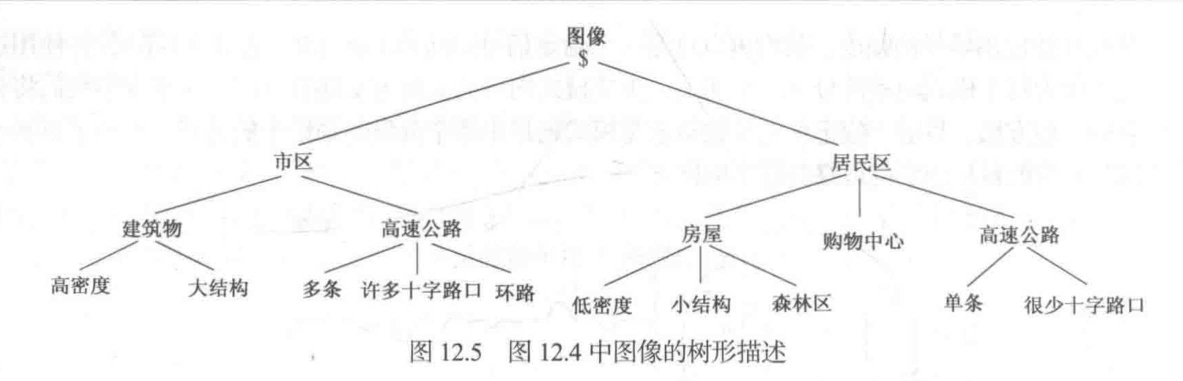


图 12.5 图 12.4 中图像的树形描述

在接下来的几节中，我们将探讨使用前几段中讨论过的技术描述的目标识别方法。

12.2 基于决策理论方法的识别

决策理论方法识别以使用决策(或判别)函数为基础。如 12.1 节中讨论的那样，令 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 表示一个 n 维模式向量。对于 W 个模式类 $\omega_1, \omega_2, \cdots, \omega_W$，决策理论模式识别的基本问题是依据如下属性来找到 W 个决策函数 $d_1(\boldsymbol{x}), d_2(\boldsymbol{x}), \cdots, d_W(\boldsymbol{x})$：如果模式 $\boldsymbol{x}$ 属于类 ω_i，则

$$d_i(\boldsymbol{x}) > d_i(\boldsymbol{x}) \qquad j=1,2,\cdots,W;\ j \neq i \tag{12.2-1}$$

换句话说，将 $\boldsymbol{x}$ 代入所有决策函数后，如果 $d_i(\boldsymbol{x})$ 得到最大值，则称未知模式 $\boldsymbol{x}$ 属于第 i 个模式类。该关系可以任意求解。

从类 ω_j 中分离出类 ω_i 的决策边界，由满足 $d_i(\boldsymbol{x})=d_j(\boldsymbol{x})$ 的 $\boldsymbol{x}$ 值给出，或等价地由

$$d_i(\boldsymbol{x})-d_i(\boldsymbol{x})=0 \tag{12.2-2}$$

的 $\boldsymbol{x}$ 值给出。

通常的做法是，使用单一函数 $d_{ij}(\boldsymbol{x})=d_i(\boldsymbol{x})-d_j(\boldsymbol{x})=0$ 来识别两类之间的决策边界。因此，对于模式类 ω_i 有 $d_{ij}(\boldsymbol{x})>0$，而对于模式类 ω_j 有 $d_{ij}(\boldsymbol{x})<0$。本节的主要目的是探讨寻找满足式(12.2-1)的决策函数的各种方法。

12.2.1 匹配

基于匹配的识别技术通过一个原型模式向量来表示每个类。根据一种预先定义的测度，将一个未知模式赋予最接近的类。最简单的方法是最小距离分类器，如其名称所示，最小距离分类器计算该未知模式与每个原型向量间的(欧氏)距离。它选择最小距离来决策。我们还将探讨一种基于相关的方法，该方法可根据图像用公式来直接表达，因此相当直观。

最小距离分类器

假设我们把每个模式类的原型定义为该类模式的平均向量：

$$\boldsymbol{m}_j=\frac{1}{N_j}\sum_{x\in\omega_j}\boldsymbol{x}_j \qquad j=1,2,\cdots,W \tag{12.2-3}$$

式中，N_j 是来自 ω_j 类模式向量的数量，求和操作对所有这些向量执行。如以前那样，W 是模式类的数量。如前所述，求未知模式向量 $\boldsymbol{x}$ 的类成员的一种方法是，将它赋给其最接近的原型类。使用欧氏距离求接近程度可将该问题简化为计算如下距离测度：

887～888

$$D_j(\boldsymbol{x})=\left\|\boldsymbol{x}-\boldsymbol{m}_j\right\| \qquad j=1,2,\cdots,W \tag{12.2-4}$$

式中，$\|\boldsymbol{a}\|=(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})^{1/2}$是欧几里得范数。然后，若$D_i(\boldsymbol{x})$是最小距离，则把$\boldsymbol{x}$赋给类$\omega_i$。也就是说，最小距离意味着该式表示最好的匹配。不难证明(见习题12.2)，选择最小距离等同于计算函数

$$d_j(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{m}_j-\frac{1}{2}\boldsymbol{m}_j^{\mathrm{T}}\boldsymbol{m}_j \qquad j=1,2,\cdots,W \tag{12.2-5}$$

并在$d_i(\boldsymbol{x})$获得最大数值时，将$\boldsymbol{x}$划归类ω_i。该式与式(12.2-1)中定义的决策函数的概念一致。

由式(12.2-2)和式(12.2-5)，对于一个最小距离分类器，类ω_i和类ω_j之间的决策边界为

$$\begin{aligned} d_{ij}(\boldsymbol{x})&=d_i(\boldsymbol{x})-d_j(\boldsymbol{x})\\ &=\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{m}_i-\boldsymbol{m}_j)-\frac{1}{2}(\boldsymbol{m}_i-\boldsymbol{m}_j)^{\mathrm{T}}(\boldsymbol{m}_i+\boldsymbol{m}_j)=0 \end{aligned} \tag{12.2-6}$$

由式(12.2-6)给出的决策面是连接$\boldsymbol{m}_i$和$\boldsymbol{m}_j$的线段的垂直等分线(见习题12.3)。$n=2$时，垂直等分线是一条直线；$n=3$时，它是一个平面；$n>3$时，称其为一个超平面。

例12.1　最小距离分类器的说明。

图12.6显示了从图12.1的鸢尾属植物例子中提取的两个模式类。两个模式类*Iris versicolor*和*Iris setosa*分别表示为ω_1和ω_2，其样本均值向量为$\boldsymbol{m}_1=(4.3,1.3)^{\mathrm{T}}$和$\boldsymbol{m}_2=(1.5,0.3)^{\mathrm{T}}$。由式(12.2-5)可知，决策函数为

$$d_1(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{m}_1-\frac{1}{2}\boldsymbol{m}_1^{\mathrm{T}}\boldsymbol{m}_1=4.3x_1+1.3x_2-10.1$$

和

$$d_2(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{m}_2-\frac{1}{2}\boldsymbol{m}_2^{\mathrm{T}}\boldsymbol{m}_2=1.5x_1+0.3x_2-1.17$$

由式(12.2-6)可知，边界方程为

$$d_{12}(\boldsymbol{x})=d_1(\boldsymbol{x})-d_2(\boldsymbol{x})=2.8x_1+1.0x_2-8.9=0$$

图12.6显示了该边界的图形(注意，轴的比例不同)。代入来自类ω_1的任何模式向量，都会得到$d_{12}(\boldsymbol{x})>0$。相反，代入来自类ω_2的任何模式向量，都会得到$d_{12}(\boldsymbol{x})<0$。换句话说，给定一个属于这两个类之一的未知模式，$d_{12}(\boldsymbol{x})$的符号将足以确定该模式的归属。

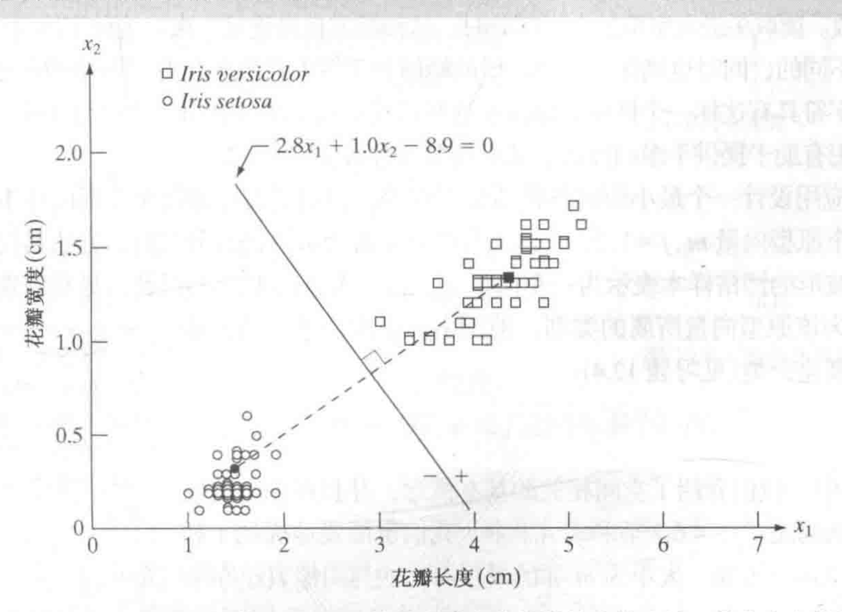


图12.6　类*Iris versicolor*和类*Iris setosa*的最小距离分类器的决策边界。黑点和方块是均值

889
~
890

实际上，就每个类的均值而言，当均值间的距离与每个类的分散度或随机性相比较大时，最小距离分类器工作得很好。在 12.2.2 节中，我们将证明：当每个类关于其均值的分布在 n 维模式空间中表现为一种球形的“超云团”形式时，最小距离分类器(在最小化错误分类的平均损失方面)会产生最佳性能。

实际中，很少同时出现较大的均值分离和相对较小的散布，除非系统设计人员能控制输入的性质。一个很好的例子就是设计用于读取固定格式字符的系统，如我们熟知的美国银行家协会的 E-13B 字型的字符集。如图 12.7 所示，这个特殊字体的字符集由 14 个字符组成，为便于读取,这些字符被特意设计在大小为 9×7 的网格中。这些字符通常使用含有很细的磁性材料的油墨来印刷。读取这些字符之前，对油墨施加了一个磁场来加重每个字符，以便简化检测。换句话说，通过人为地突出每个字符的关键特征，解决了分割问题。

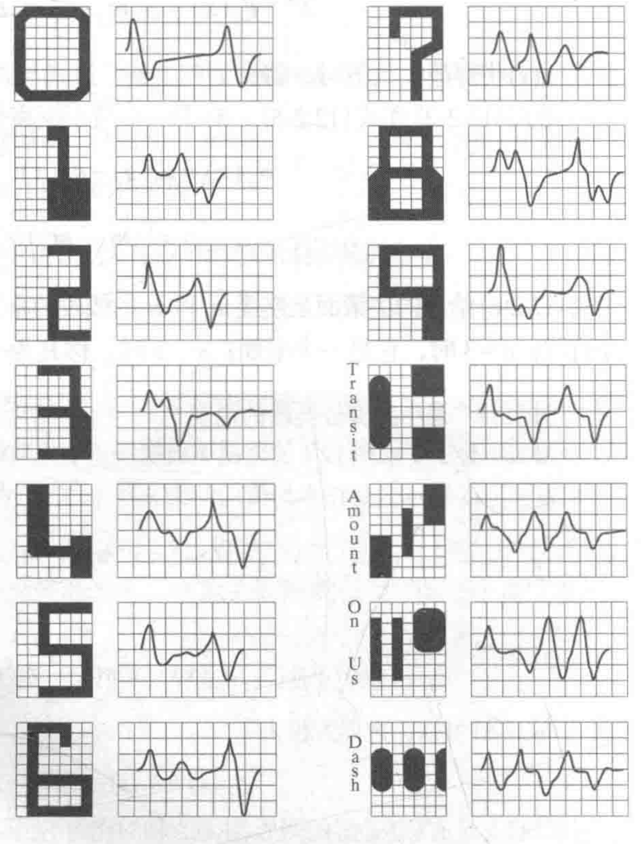


图 12.7 美国银行家协会 E-138 字符集和对应的波形

通常使用单缝读取头来水平扫描字符，读取头与字符相比要窄一些、长一些。当读取头横向移过一个字符时,会产生一个一维电信号(一个信号),该电信号与读取头下方字符面积的增大率或减小率成正比。例如，考虑与图12.7 中数字 0 相关联的波形。读取头从左向右移动时，读取头看到的字符面积开始增大,产生一个正导数(一个正变化率)。读取头开始远离 0 的左半部时，读取头下方的字符面积开始减少，产生一个负导数。读取头位于该字符的中间区域时，字符面积几乎保持不变，产生一个零导数。读取头进入字符的右半部分时，这种模式自身重复一次。这种字体设计确保了各个字符的波形是完全不同的，同时也确保了每个波形的峰值和零值大致出现在背景网格的垂直线上，如图 12.7 所示。E-13B 字符具有这样一个性质，即仅在这些点处对波形取样就可产生进行正确分类的足够信息。磁性油墨的使用有助于提供干净的波形，从而使分散性最小。

针对这一应用设计一个最小距离分类器相当简单。我们只需存储每个波形的样本值，然后将每组样本表示为一个原型向量 $\boldsymbol{m}_j, j = 1, 2, \cdots, 14$。当对一个未知字符进行分类时，方法是使用刚才描述的方式扫描它，将波形的网格样本表示为一个向量 $\boldsymbol{x}$，然后依据式(12.2-5)得到的最高值选择原型向量的类别。该类别即为该原型向量所属的类别。使用由一组电阻器组成的模拟电路可实现高速分类(见习题 12.4)。

> 当函数不同时，我们应将相关正式地称为互相关，而当函数相同时，则应将相关称为自相关。但是，当一个给定应用中两个函数相同或不同很明确时，通常使用相关这一通用术语。

相关匹配

在3.4.2 节中，我们介绍了空间相关的基本概念，并且在该节将它广泛地用于空间滤波。在 4.6.7 节和表 4.3 中，我们也简要地提到了相关定理。由式(3.4-1)可知，大小为 $m \times n$ 的模板 $w(x, y)$与图像 $f(x, y)$的相关可表示为

$$c(x,y) = \sum_s \sum_s w(s,t) f(x+s, y+t) \tag{12.2-7a}$$

890~892

其中，求和的上下限取 w 和 f 的共同范围。对变量 x 和 y 的所有偏移值计算该式，以便 w 的所有元素访问 f 的每个像素，这里假设 f 大于 w。就像空间卷积通过卷积定理与函数的傅里叶变换相联系那样，空间相关通过相关定理与函数的变换相联系：

$$f(x,y)☆w((x,y)\Leftrightarrow F^*(u,v)W(u,v) \tag{12.2-7b}$$

式中，☆表示空间相关，F^* 是 F 的复共轭。表 4.3 中说明的相关定理的另一半对当前讨论而言并不重要。式(12.2-7b)是傅里叶变换对，除了我们使用函数的一个复共轭之外，其解释与式(4.6-24)是一样的。式(12.2-7b)的傅里叶反变换得到一个类似于式(4.6-23)的二维循环相关，并且 4.6.6 节中讨论的卷积填充问题同样适用于相关。

我们并不打算详述前述公式，因为它们对 f 和 w 的尺度变化很敏感。作为替代，我们使用如下的归一化相关系数：

$$\gamma(x,y)=\frac{\sum_s\sum_t[w(s,t)-\overline{w}]\left[f(x+s,y+t)-\overline{f}_{xy}\right]}{\left\{\sum_s\sum_t[w(s,t)-\overline{w}]^2\sum_s\sum_t\left[f(x+s,y+t)-\overline{f}_{xy}\right]^2\right\}^{\frac{1}{2}}} \tag{12.2-8}$$

其中，求和的上下限取 w 和 f 的共同范围，$\overline{w}$ 是模板的平均值(只计算一次)，$\overline{f}_{xy}$ 是 f 中与 w 重合区域的平均值。通常，我们将 w 称为模板，而将相关称为模板匹配。可以证明(见习题 12.7)，$\gamma(x,y)$的值域为$[-1, 1]$，因而 f 和 w 的幅度变化是归一化的。当归一化的 w 和 f 中对应的归一化区域相同时，$\gamma(x,y)$出现最大值。这说明了最大相关(即最好可能的匹配)。当两个归一化函数在式(12.2-8)的意义下表现出最小相似性时，会出现最小值。相关系数不能用傅里叶变换来计算，因为该式中存在非线性项(除法和平方)。

回顾 3.4.2 节有助于我们了解空间相关的机理。

图 12.8 说明了刚才讨论的步骤的机理。如 3.4.2 节中解释的那样，当 w 的中心位于 f 的边界上时，围绕 f 的边界需要进行填充(在模板匹配中，当模板的中心越过图像的边界时，相关的值通常并不重要，因此填充被限制为模板宽度的一半)。为表示方便，我们通常只关心奇数大小的模板。

图 12.8 显示了一个大小为 $m\times n$ 的模板，其中心位于任意位置(x, y)。在这一点的相关可以用式(12.2-8)得到。然后，模板的中心移到一个相邻位置，重复该过程。通过移动模板的中心(即增大 x 和 y)，以便 w 的中心访问 f 中的每个像素，可得到所有的相关系数 $\gamma(x,y)$。在这一过程的最后，我们寻找 $\gamma(x,y)$中的最大值，进而找到出现最好匹配的位置。在 $\gamma(x,y)$中可能会有多个位置出现最大值，此时表明 w 和 f 之间有多个匹配。

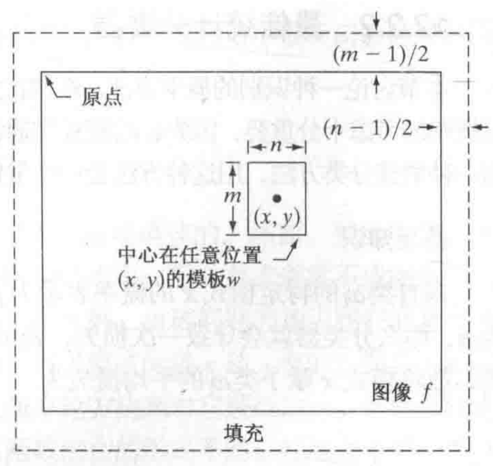


图 12.8　模板匹配的机理

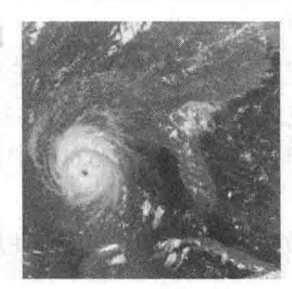
例 12.2　用相关进行匹配。

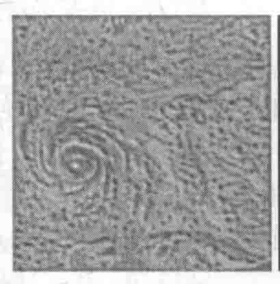
图12.9(a)显示了一幅大小为 913×913 的“安德鲁”飓风卫星图像，其中暴风眼清晰可见。作为相关的一个例子，我们希望找到图12.9(b)中的模板在图(a)中最好匹配的位置，模板是暴风眼的一幅大小为 31×31 的子图像。图 12.9(c)显示了由式(12.2-8)计算的相关系数。由于填充(见图 12.8)，该图像的原始大小为 943×943 像素，但为了显示的目的，我们把它裁剪到原始图像大小。在该图像中，灰度与相关值成

892 ~ 893

正比，并且所有的负相关都截为0(黑)，以简化该图像的视觉分析。相关图像的最亮点在接近飓风眼的位置清晰可见。图12.9(d)将最大相关显示为一个白点(此时，存在一个最大值为1的唯一匹配)，我们看到它与图12.9(a)中的飓风眼位置紧密对应。

a b c d





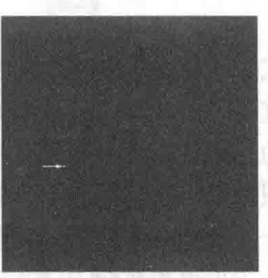


图12.9 (a)飓风“安德鲁”的卫星图像，摄于1992年8月24日；(b) 飓风眼的模板；(c)显示为图像的相关系数(注意最亮点)；(d)匹配最好的位置。该点是单个像素，但其已被放大，以便于查看(原图像由 NOAA 提供)

前述讨论表明，可以对被处理的函数的灰度值变化进行归一化相关。对尺寸和旋转归一化是一个更复杂的问题。尺寸归一化涉及空间缩放，正如2.6.5节和4.5.4节中解释的那样，即图像重采样。为了使重采样有意义，必须知道被重新缩放的图像的大小。在某些情况下，这可能变成一个很困难的问题，除非是在有空间线索可用的情况下。例如，在遥感应用中，如果已知成像传感器的观察几何(典型情况)，假设视角固定，那么知道传感器相对于成像区域的高度将足以归一化图像尺寸。类似地，旋转归一化要求已知图像被旋转的角度。这再一次要求空间线索。在刚才给出的遥感例子中，飞行方向可能足以把遥感图像旋转到一个标准方向。在没有约束条件的情形下，尺寸和旋转归一化会成为相当具有挑战性的任务，因为它要求自动检测可作为空间线索的图像特性(如11章讨论的那样)。

12.2.2 最佳统计分类器

本节讨论一种识别的概率方法。就像在大多数处理测度和判读物理事件的领域中那样，在模式识别中对概率的考虑十分重要，因为此时模式类通常会随机生成。如下面的讨论所示，在平均意义上有可能推导出一种最佳分类方法，用这种方法会产生最低的错误分类的概率(见习题12.10)。

基础知识

来自类ω_i的特定模式$\boldsymbol{x}$的概率表示为$p(\omega_i/\boldsymbol{x})$。如果模式分类器判断$\boldsymbol{x}$来自类$\omega_j$，而实际上它来自类$\omega_i$，那么分类器就会导致一次损失，表示为$L_{ij}$。由于模式$\boldsymbol{x}$可能属于所考虑的$W$个类中的任何一个类，故将模式$\boldsymbol{x}$赋予类$\omega_j$的平均损失为

$$r_j(\boldsymbol{x}) = \sum_{k=1}^{W} L_{kj}\, p(\omega_k / \boldsymbol{x}) \tag{12.2-9}$$

该式在决策理论术语中通常称为条件平均风险或损失。

> 关于概率论的简要回顾，请读者查阅本书的网站。

由基础概率论可知，$p(A/B) = [p(A)p(B/A)]/p(B)$。使用该式，我们可将式(12.2-9)写为

$$r_j(\boldsymbol{x}) = \frac{1}{p(\boldsymbol{x})} \sum_{k=1}^{W} L_{kj}\, p(\boldsymbol{x} / \omega_k) P(\omega_k) \tag{12.2-10}$$

893 ~ 895

式中，$p(\boldsymbol{x}/\omega_k)$是来自类$\omega_k$的模式的概率密度函数，$P(\omega_k)$是类$\omega_k$出现的概率(有时这些概率称为先验概率)。由于$1/p(\boldsymbol{x})$为正，且对所有的$r_j(\boldsymbol{x}), j = 1, 2, \cdots, W$都是如此，故可将它从式(12.2-10)中忽

略而不影响这些函数从最小值到最大值的相对顺序。然后，平均损失的表达式就简化为

$$r_j(\boldsymbol{x}) = \sum_{k=1}^{W} L_{kj} p(\boldsymbol{x}/\omega_k)P(\omega_k) \tag{12.2-11}$$

分类器有 W 个可能的类，任何给定的未知模式可从这些类中选择。如果分类器为每个模式 $\boldsymbol{x}$ 计算 $r_1(\boldsymbol{x}), r_2(\boldsymbol{x}), \cdots, r_W(\boldsymbol{x})$，并将该模式以最低损失赋给相应的类，则关于所有决策的总体平均损失将是最低的。这种将总体平均损失降至最低的分类器称为贝叶斯分类器。因此，如果 $r_i(\boldsymbol{x}) < r_j(\boldsymbol{x}), j = 1, 2, \cdots, W$ 且 $j \neq i$，那么贝叶斯分类器将未知模式 $\boldsymbol{x}$ 赋给类 ω_i。换句话说，如果对所有的 j 且 $j \neq i$ 有

$$r_j(\boldsymbol{x}) = \sum_{k=1}^{W} L_{kj} p(\boldsymbol{x}/\omega_k)P(\omega_k) < \sum_{q=1}^{W} L_{qj} p(\boldsymbol{x}/\omega_q)P(\omega_q) \tag{12.2-12}$$

那么 $\boldsymbol{x}$ 将赋给类 ω_i。正确决策的损失通常被赋予零值，而不正确决策的损失通常被赋予相同的非零值(譬如值 1)。在这些条件下，损失函数变为

$$L_{ij} = 1 - \delta_{ij} \tag{12.2-13}$$

式中，$i = j$ 时 $\delta_{ij} = 1$，$i \neq j$ 时 $\delta_{ij} = 0$。式(12.2-13)表明，不正确决策的损失是 1，正确决策的损失是 0。将式(12.2-13)代入式(12.2-11)得到

$$r_j(\boldsymbol{x}) = \sum_{k=1}^{W} (1 - \delta_{kj}) p(\boldsymbol{x}/\omega_k)P(\omega_k) = P(\boldsymbol{x}) - p(\boldsymbol{x}/\omega_j)P(\omega_j) \tag{12.2-14}$$

如果对所有的 $j \neq i$ 有

$$p(\boldsymbol{x}) - p(\boldsymbol{x}/\omega_i)P(\omega_i) < P(\boldsymbol{x}) - p(\boldsymbol{x}/\omega_j)P(\omega_j) \tag{12.2-15}$$

或等价地有

$$p(\boldsymbol{x}/\omega_j)P(\omega_j) > p(\boldsymbol{x}/\omega_j)P(\omega_j) \quad j = 1, 2, \cdots, W;\ j \neq i \tag{12.2.16}$$

那么贝叶斯分类器将模式 $\boldsymbol{x}$ 赋给类 ω_i。参考推导出式(12.2-1)的讨论，我们知道 0-1 损失函数的贝叶斯分类器不过是如下形式的决策函数的计算：

$$d_j(\boldsymbol{x}) = p(\boldsymbol{x}/\omega_j)P(\omega_j) \quad j = 1, 2, \cdots, W \tag{12.2-17}$$

式中，模式向量 $\boldsymbol{x}$ 赋给其决策函数取得最大值的类。

式(12.2-7)给出的决策函数在最小化错误分类中的平均损失方面是最佳的。然而，要保持这一最佳性，就必须知道每个类中的模式的概率密度函数及每个类出现的概率。后一要求通常不成问题。例如，如果所有的类等概率出现，那么 $p(\omega_j) = 1/W$。即使该条件不正确，由该问题的知识我们通常也可以推出这些概率。概率密度函数 $p(\boldsymbol{x}/\omega_j)$的估计则是另一回事。如果模式向量 $\boldsymbol{x}$ 是 n 维的，那么 $p(\boldsymbol{x}/\omega_j)$ 是一个 n 元函数，如果该函数的形式未知，就需要使用多元概率论方法来对它进行估计。这些方法在实际应用中很困难，代表每一类的模式数目不大或概率密度函数潜在的形式不能很好地表示时，尤其如此。由于这些原因，使用贝叶斯分类器时，通常假设对各种密度函数有一个解析表达式，且来自每个类的样本模式有一个必需的参数估计。目前，$p(\boldsymbol{x}/\omega_j)$的最为通用的假设形式是高斯概率密度函数。这种假设与真实情况越接近，贝叶斯分类器方法在分类中就越能接近最小平均损失。

高斯模式类的贝叶斯分类器

我们先考虑一个一维问题，该问题包含有由高斯密度决定的两个模式类(W=2)，这两个模式类的均值分别为 m_1 和 m_2，标准差分别为 σ_1 和 σ_2。由式(12.2-17)可知，贝叶斯决策函数的形式为

895 ~ 896

$$d_j(x)=p(x/\omega_j)P(\omega_j)=\frac{1}{\sqrt{2\pi}\sigma_j}\mathrm{e}^{-\frac{(x-m_j)^2}{2\sigma_j^2}}P(\omega_j),\qquad j=1,2 \tag{12.2-18}$$

像 10.3.3 节中提及的那样，一元贝叶斯分类器是最优阈值处理函数的事实，请参阅本节末尾的注释。

其中，这些模式现在是标量，由 x 表示。图 12.10 显示了这两类的概率密度函数曲线。两类之间的边界是一个单点，它由满足 $d_1(x_0)=d_2(x_0)$ 的 x_0 表示。如果两个类出现的概率相同，则 $P(\omega_1)=P(\omega_2)=1/2$，且决策边界是满足 $p(x_0/\omega_1)=p(x_0/\omega_2)$的 x_0 值。如图 12.10 所示，该点是两个概率密度函数的交点。任何处于 x_0 点右侧的模式(点)都被划归类ω_1。同样，任何处于 x_0 点左侧的模式(点)都被划归类ω_2。当两个类出现的概率不相等时，如果类ω_1 出现的概率大，则点 x_0 左移；反之，如果类ω_2 出现的概率大，则点 x_0 右移。这是我们期望的结果，因为分类器总是试图将错误分类的损失降至最低。例如，在极端情形下，如果类ω_2 从不出现，那么分类器总是将所有模式划归类ω_1(即点 x_0 移向负无穷大)，故从不会出现错误。

在 n 维情形下，第 j 个模式类中的向量的高斯密度为

$$p(\boldsymbol{x}/\omega_j)=\frac{1}{(2\pi)^{n/2}\left|\boldsymbol{C}_j\right|^{1/2}}\mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}_j)^{\mathrm{T}}\boldsymbol{C}_j^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)} \tag{12.2-19}$$

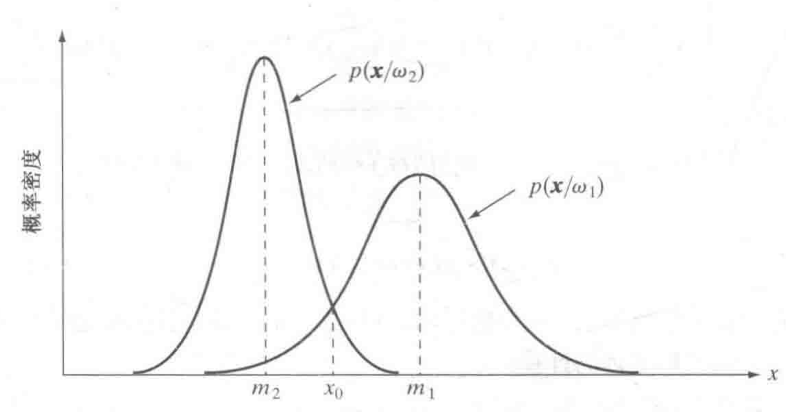


图 12.10　两个一维模式类的概率密度函数。如果两个类出现的概率相等，那么所示的点 x_0 就是决策边界

其中，每个密度完全由其均值向量 $\boldsymbol{m}_j$ 和协方差矩阵 $\boldsymbol{C}_j$ 确定，均值向量和协方差矩阵定义如下：

$$\boldsymbol{m}_j=E_j\{\boldsymbol{x}\} \tag{12.2-20}$$

和

$$\boldsymbol{C}_j=E_j\left\{(\boldsymbol{x}-\boldsymbol{m}_j)(\boldsymbol{x}-\boldsymbol{m}_j)^{\mathrm{T}}\right\} \tag{12.2-21}$$

式中，$E_j\{\cdot\}$ 表示类ω_j 的模式的自变量的期望值。在式(12.2-19)中，n 是模式向量的维度，$|\boldsymbol{C}_j|$是矩阵 $\boldsymbol{C}_j$ 的行列式。使用所讨论量的平均值来近似期望值 E_j，可得到均值向量和协方差矩阵的估计，即

$$\boldsymbol{m}_j=\frac{1}{N_j}\sum_{\boldsymbol{x}\in\omega_j}\boldsymbol{x} \tag{12.2-22}$$

和

$$\boldsymbol{C}_j=\frac{1}{N_j}\sum_{\boldsymbol{x}\in\omega_j}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}-\boldsymbol{m}_j\boldsymbol{m}_j^{\mathrm{T}} \tag{12.2-23}$$

897 ~ 898

式中，N_j 是来自类ω_j 的模式向量的数量，求和操作对所有这些向量执行。在本节后面，我们将给出一个如何使用这两个表达式的例子。

关于向量和矩阵的简要回顾，请查阅本书的网站。

协方差矩阵是对称的和半正定的。如11.4节中解释的那样，对角线元素 c_{kk} 是模式向量的第 k 个元素的方差，非对角线元素 c_{jk} 是 x_j 和 x_k 的协方差。当协方差矩阵的非对角线元素均为零时，多元高斯密度函数就简化为 $\boldsymbol{x}$ 的每个元素的单变量的高斯密度的乘积。当向量元素 x_j 和 x_k 不相关时，会出现这种情况。

根据式(12.2-17)，类 ω_j 的贝叶斯决策函数为 $d_j(\boldsymbol{x}) = p(\boldsymbol{x}/\omega_j)P(\omega_j)$。然而，由于高斯密度函数的指数形式，用这个决策函数的自然对数形式更为方便。换句话说，我们可以使用如下形式：

$$d_j(\boldsymbol{x}) = \ln\left[p(\boldsymbol{x}/\omega_j)P(\omega_j)\right] = \ln\left[p(\boldsymbol{x}/\omega_j) + \ln P(\omega_j)\right] \tag{12.2-24}$$

从分类性能方面看，该表达式等同于式(12.2-17)，因为对数函数是单调递增函数。换句话说，式(12.2-17)中的决策函数的数值顺序与式(12.2-24)中的数值顺序是相同的。将式(12.2-19)代入式(12.2-24)中可得

$$d_j(\boldsymbol{x}) = \ln P(\omega_j) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\left|\boldsymbol{C}_j\right| - \frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{m}_j)^{\mathrm{T}}\boldsymbol{C}_j^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)\right] \tag{12.2-25}$$

项 $(n/2)\ln 2\pi$ 对所有的类都是相同的，因此可以从式(12.2-25)中消去，消去该项后的式(12.2-25)变为

$$d_j(\boldsymbol{x}) = \ln P(\omega_j) - \frac{1}{2}\ln\left|\boldsymbol{C}_j\right| - \frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{m}_j)^{\mathrm{T}}\boldsymbol{C}_j^{-1}(\boldsymbol{x}-\boldsymbol{m}_j)\right] \tag{12.2-26}$$

式中 $j = 1, 2, \cdots, W$。式(12.2-26)表示高斯模式类在0-1损失函数条件下的贝叶斯决策函数。

式(12.2-26)中的决策函数是超二次曲面(n 维空间中的二次函数)，因为出现在该式中的 $\boldsymbol{x}$ 的分量的各项没有高于二次的。毫无疑问，高斯模式的贝叶斯分类器的作用至多是将一个通用二阶决策面放到两个模式类之间。然而，如果模式总体确实是高斯的，则在分类中不会有其他曲面所产生的损失小于平均损失。

如果所有的协方差矩阵都相等，则 $\boldsymbol{C}_j = \boldsymbol{C}, j = 1, 2, \cdots, W$。展开式(12.2-26)并消去所有独立于 j 的项，可得

$$d_j(\boldsymbol{x}) = \ln P(\omega_j) + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{m}_j - \frac{1}{2}\boldsymbol{m}_j^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{m}_j \tag{12.2-27}$$

这是一个线性决策函数(超平面)，其中 $j = 1, 2, \cdots, W$。

另外，如果 $\boldsymbol{C} = \boldsymbol{I}$，这里 $\boldsymbol{I}$ 是单位矩阵，且 $P(\omega_j) = 1/W, j = 1, 2, \cdots, W$，则有

$$d_j(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{m}_j - \frac{1}{2}\boldsymbol{m}_j^{\mathrm{T}}\boldsymbol{m}_j \qquad j = 1, 2, \cdots, W \tag{12.2-28}$$

这些是最小距离分类器的决策函数，就像式(12.2-5)中给出的一样。如果：(1)这些模式类都是高斯的，(2)所有协方差矩阵都等于单位矩阵，并且(3)所有类出现的概率相等，则在贝叶斯意义上最小距离分类器是最佳的。满足这些条件的高斯模式类是 n 维空间中外形相同的球状云团(称为超球面)。最小距离分类器在每对类之间建立一个超平面，超平面有如下特性，它是连接这对超球面的中心的垂直平分线。在两个维度上，这些类组成多个圆形区域，并且边界变为连接每对圆的中心的直线的垂直平分线。

例 12.3　三维模式的贝叶斯分类器。

图12.11显示了两个模式类在三维空间中的一个简单排列。我们用这些模式来说明实现贝叶斯分类器的机理，假设每个类的模式都是来自一个高斯分布的样本。

对图12.11中的模式应用式(12.2-22)，得到

$$\boldsymbol{m}_1 = \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix} \quad 和 \quad \boldsymbol{m}_2 = \frac{1}{4}\begin{bmatrix}1\\3\\3\end{bmatrix}$$

898～899

类似地，对两个模式类依次应用式(12.2-23)会得到两个协方差矩阵，此时这两个矩阵是相等的：

$$C_1 = C_2 = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

因为协方差矩阵相等，故贝叶斯决策函数由式(12.2-27)给出。若假设 $P(\omega_1) = P(\omega_2) = 1/2$，则应用式(12.2-28)可得

$$d_j(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{m}_j - \frac{1}{2}\boldsymbol{m}_j^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{m}_j$$

式中，

$$\boldsymbol{C}^{-1} = \begin{bmatrix} 8 & -4 & -4 \\ -4 & 8 & 4 \\ -4 & 4 & 8 \end{bmatrix}$$

对 $d_j(\boldsymbol{x})$执行向量矩阵展开，给出决策函数如下：

$$d_1(\boldsymbol{x}) = 4x_1 - 1.5 \quad 和 \quad d_2(\boldsymbol{x}) = -4x_1 + 8x_2 + 8x_3 - 5.5$$

分隔两个类的决策面为

$$d_1(\boldsymbol{x}) - d_2(\boldsymbol{x}) = 8x_1 - 8x_2 - 8x_3 + 4 = 0$$

图 12.11 显示了该决策面的剖面，我们注意到两个类已被有效地分开。

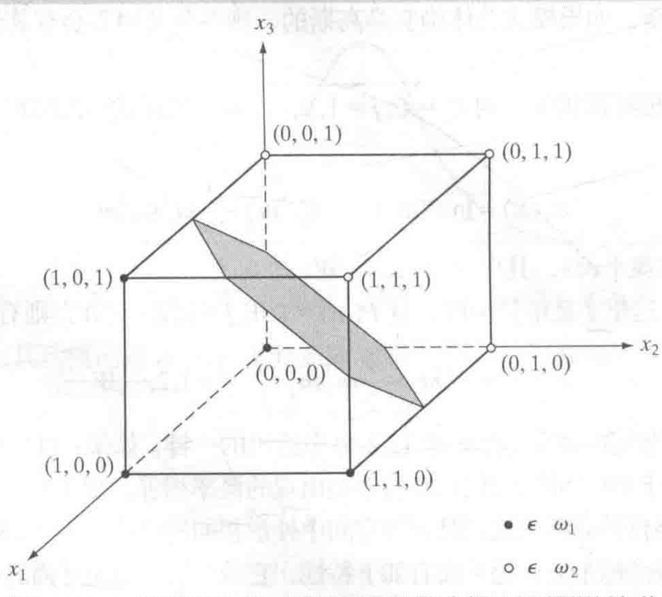


图 12.11　两个简单的模式类及其贝叶斯决策边界(阴影所示)

贝叶斯分类器方法最成功的应用之一是，对由飞行器、卫星或空间站上携带的多光谱扫描器生成的遥感图像进行分类。对这些平台上产生的大量图像数据使得图像自动分类和分析，已成为遥感的重要任务。遥感应用相当广泛，包括土地利用调查，农作物产量调查，农作物灾害检测，森林、空气质量和水质监测，地质研究，天气预报，以及其他许多对环境有重要意义的领域。下例显示了一个典型应用。

例 12.4　使用贝叶斯分类器对多光谱数据进行分类。

如 1.3.4 节和 11.4 节讨论的那样，多光谱扫描器响应于所选电磁能谱的波段，例如 0.45~0.52 μm, 0.52~0.60 μm, 0.63~0.69 μm 和 0.76~0.90 μm 波段。这些范围分别处于可见蓝光、可见绿光、可见红光和近

899 ~ 901

红外波段。按这种方式扫描一个区域，会生成该区域的 4 幅数字图像，每个波段一幅。如果这些图像已在空间上配准（实践中满足的一个条件），那么可将这些图像想象为堆叠在一起，如图12.12 所示。这样，就像我们在 11.4 节中所做的那样，地面上的每个点都可以由形如 $\boldsymbol{x}=(x_1,x_2,x_3,x_4)^{\mathrm{T}}$ 的一个四元素模式向量表示，其中 x_1 为蓝光图像，x_2 为绿光图像，等等。如果这些图像的大小都是512×512像素，那么这 4 幅多光谱图像的每种堆叠可由266 144个四维模式向量表示。像前面说明的那样，高斯模式的贝叶斯分类器要求估计每个类的均值向量和协方差矩阵。在遥感应用中，这些估计是通过从每个感兴趣区域中收集其类别已知的多光谱数据得到的。然后，就像在例 12.3 中那样，结果向量用于估计所需的均值向量和协方差矩阵。

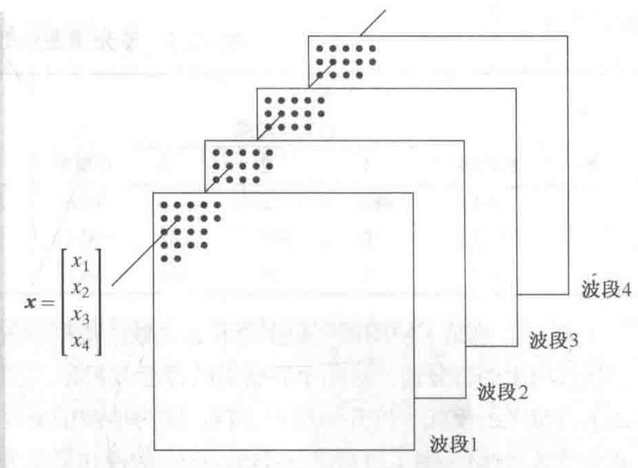


图 12.12 由多光谱扫描器生成的 4 幅数字图像经像素配准后，所形成的模式向量

图 12.13(a)到(d)是华盛顿特区的 4 幅大小为 512×512 的多光谱图像，这些图像是在前段中提及的波段范围拍摄的。我们的兴趣是把由这些图像包含的区域中的像素分类到 3 个模式类之一：水体、市区和植被。图 12.13(e)中的模板已叠加到这些图像上，以提取这 3 个类的典型样本。样本的一半用于训练（即估计均值向量和协方差矩阵），另一半用于独立测试，以评估初始分类器的性能。在无约束条件的多光谱数据分类中，由于通常并不知道先验概率 $P(\omega_i)$，所在我们在这里假设它们是相等的，即 $P(\omega_i)=1/3, i=1,2,3$。

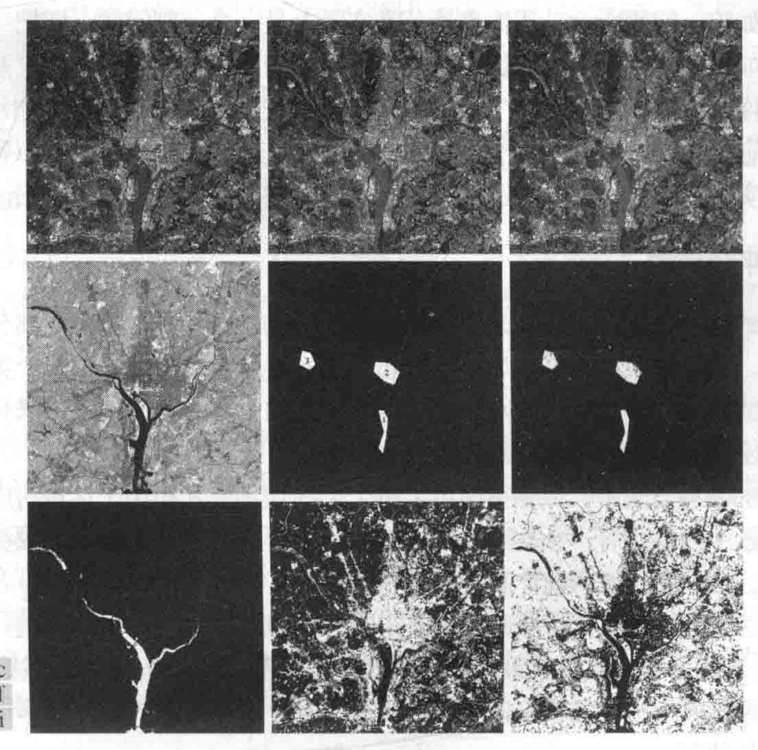


图 12.13 多光谱数据的贝叶斯分类：(a)~(d)可见蓝光、可见绿光、可见红光和近红外波长图像；(e)显示(1)水体、(2)市区和(3)植被的样本区域的模板；(f)分类结果。黑点表示未正确分类的点，其他(白)点是正确分类的点；(g)分类为水体的所有图像像素(白色)；(h)分类为市区的所有图像像素(白色)；(i)分类为植被的所有图像像素(白色)

901
~
903

表 12.1 多光谱图像数据的贝叶斯分类

训练模式						无关模式					
		已划分的类			%			已划分的类			%
类	样本数	1	2	3	正确率	类	样本数	1	2	3	正确率
1	484	482	2	0	99.6	1	483	478	3	2	98.9
2	933	0	885	48	94.9	2	932	0	880	52	94.4
3	483	0	19	464	96.1	3	482	0	16	466	96.7

表 12.1 总结了使用训练数据集和独立数据集得到的识别结果。正确识别的训练模式向量的百分比和独立模式向量的百分比，对两个数据集而言近似相等，这表明了参数估计的稳定性。两种情形下的最大误差是来自市区的模式。这并未出人意料，因为植被也出现在那里(注意，在植被区域或市区没有模式被错误地分类为水体)。图 12.13(f) 中将错误分类的模式显示为黑点，而将正确分类的模式显示为白点。在区域 1 中看不到黑点，因为 7 个错误分类的点非常靠近白色区域的边界。

图 12.13(g) 到 (i) 更为有趣。这里，我们使用从训练数据中获取的均值向量和协方差矩阵，将所有图像像素分为了三个类别之一。图 12.13(g) 以白色显示了分类为水体的所有像素。未被分类为水体的像素使用黑色显示。我们看到，贝叶斯分类器对确定哪一部分是水体的工作做得很出色。图 12.13(h) 以白色显示了市区发展的所有像素；系统出色地识别了市区特性，如桥梁和高速公路。图 12.13(i) 显示了被分类为植被的像素。图 12.13(h) 表明中心城区的白色像素高度集中，其密度与到图像中心的距离为函数递减。图 12.13(i) 显示了相反的效果，表明图像的中心植被很少，而在城市开发区植被最多。

我们在 10.3.3 节开始时曾提到过，阈值处理可视为贝叶斯问题，它最佳地将模式赋给两个类或多个类。事实上，如前一问题所示，逐像素地分类实际上是一个分割问题，即把一幅图像分成两个或多个可能的区域。如果仅使用一个单变量，则式(12.2-17)变成一个最佳函数，它基于一幅图像的像素的灰度来分割该图像，就像我们在 10.3 节中所做的那样。记住，最佳性要求已知每个类的概率密度函数和先验概率。如前面提及的那样，估计这些密度并不容易。如果必须做出假设(譬如假设为高斯密度)，则在分割中所能实现的最佳程度，与这些假设接近真实情况的程度成正比。

12.2.3 神经网络

前两节讨论的方法使用样本模式来估计每个模式类的统计参数。最小距离分类器完全由每个类的均值向量来确定。类似地，对于总体为高斯分布的贝叶斯分类器，完全由每个类的均值向量和协方差矩阵确定。用于估计这些参数的(已知其所属的类)模式通常称为训练模式，来自每个类的这样一组模式称为训练集。使用训练集得到决策函数的过程称为学习或训练。

在刚才讨论的两种方法中，训练很简单。每个类的训练模式被用于计算对应于该类的决策函数的参数。估计出问题中的参数后，分类器的结构就被固定，且其最终的性能将取决于实际模式总体是否满足所用分类方法推导中所做的统计假设。

一个问题中的模式类的统计特性通常是未知的，或者是无法估计的(回想前一节中关于多元统计的难点的简单讨论)。实际上，此类决策理论问题最好由直接通过训练生成所需决策函数的方法来处理。然后，没有必要再做关于基本的概率密度函数或关于所考虑模式类的其他概率信息的假设。本节讨论满足这一准则的各种方法。

背景知识



以下内容的本质是使用大量的基本非线性计算单元(称为神经元)，这些单元以网络的形式进行组织，就像大脑中互连的神经元那样。得到的模型有各种各样的名称，包括神经网络、神经计算机、

并行分布式处理(PDP)模型、神经形态学系统、分层自适应网络和连接模型。这里，我们使用神经网络这一名称，或简称为神经网。我们使用这些网络作为工具，通过对模式的训练集的逐次描述，自适应地导出判别函数的系数。

神经网络受到关注可以追溯至 20 世纪 40 年代早期，详见 McCulloch and Pitts[1943]。他们提出了神经元模型，该模型用二进制门限器件和包括 0-1 与 1-0 的神经状态突变的随机算法，作为神经系统建模的基础。Hebb[1949]的后续工作以数学模型为基础，该模型试图通过增强或联合来掌握学习这一概念。

20 世纪 50 年代中期至 60 年代早期，Rosenblatt[1959, 1962]发明学习机引起了模式识别理论领域研究人员和从业者的巨大兴趣。人们对这些称为*感知机*的机器如此关注的原因是，数学证明的进展显示：使用线性可分的训练集(即由一个超平面分隔的训练集)进行训练时，这些感知机会在有限数量的迭代步骤后收敛到一个解。采用超平面系数形式的这个解，能正确地分离由该训练集的模式表示的类。

遗憾的是，人们很快就对看起来有良好理论基础的学习模型大失所望。当时，基本的感知机及其推广还很简单，不足以解决多数有实际意义的模式识别任务。后来，通过研究多层这样的器件，人们试图扩展类感知机的能力，尽管它们在概念上很吸引人，但都缺乏原型感知机那样有效的训练算法。Nilsson[1965]总结了 20 世纪 60 年代中期学习机领域的状况。几年后，Minsky andPapert[1969]提出了类感知机机器的局限性这一令人气馁的分析。这一看法一直持续到 20 世纪 80 年代中期，佐证文献见 Simon[1986]。在这篇最初于 1984 年以法文发表的文章中，Simon 以“一个神话的诞生和死亡”为题驳斥了感知机。

Rumelhart, Hinton and Williams[1986]对多层感知机的新训练算法的研究，使得事情出现了转机。他们的基本方法(通常称为*反向传播方式学习的通用* delta *法则*)，为多层机器提供了一种有效的训练方式。尽管不能像证明单层感知机那样证明这种训练算法也收敛到一个解，但通用 delta 法则已成功地解决了大量的实际问题。这一成功使得多层类感知机已成为当前所用神经网络的主要模型之一。

两个模式类的感知机

在这种最基本的形式中，感知机学习一个线性决策函数，该决策函数对分两个线性可分的训练集。图 12.14(a)显示了两个模式类的感知机模型。这个基本装置的响应基于其输入的加权和，即

$$d(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i + w_{n+1} \tag{12.2-29}$$

这是一个与模式向量的分量有关的线性决策函数。称为*权重*的系数 $w_i, i = 1, 2, \cdots, n, n+1$ 在对输入求和前，对这些输入进行修正，并馈送到阈值单元中。在这一意义上，权重类似于人类神经系统中的神经突触。将求和连接的输出映射为该装置的最终输出的函数，有时称为*激活函数*。

当 $d(\boldsymbol{x}) > 0$ 时，阈值单元使感知机的输出为+1，这表明模式 $\boldsymbol{x}$ 被识别为属于类 ω_1。当 $d(\boldsymbol{x}) < 0$ 时，情形正好相反。这种操作模式与之前的注释为两个类使用单个决策函数的式(12.2-2)是一致的。当 $d(\boldsymbol{x})=0$ 时，$\boldsymbol{x}$ 位于分隔两个模式类的决策面上，这给出了一个不能确定的条件。由感知机实现的决策边界是通过令式(12.2-29)等于零得到的：

$$d(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i + w_{n+1} = 0 \tag{12.2-30}$$

或



$$w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_{n+1} = 0 \tag{12.2-31}$$

这是在 n 维模式空间中的一个超平面方程。几何上，前 n 个系数确定超平面的方向，而最后一个系数 w_{n+1} 与从原点到超平面的垂直距离成正比。因此，若 $w_{n+1}=0$，则超平面通过模式空间的原点。类似地，若 $w_j=0$，则超平面平行于 x_j 轴。

图 12.14(a)中阈值单元的输出取决于 $d(\boldsymbol{x})$的符号。替代测试整个函数来确定它是正还是负，我们可以对 w_{n+1} 项检验式(12.2-29)的求和部分，此时系统的输出是

$$O=\begin{cases}+1, & \sum_{i=1}^{n} w_1 x_1 > -w_{n+1}\\ -1, & \sum_{i=1}^{n} w_1 x_1 < -w_{n+1}\end{cases} \tag{12.2-32}$$

该实现等同于图 12.14(a)并显示在图 12.14(b)中，唯一的不同是，阈值函数被量 $-w_{n+1}$ 取代，且常量单位输入不再出现。在本节后面讨论多层神经网络的实现时，我们会回顾这两个公式的等价性。

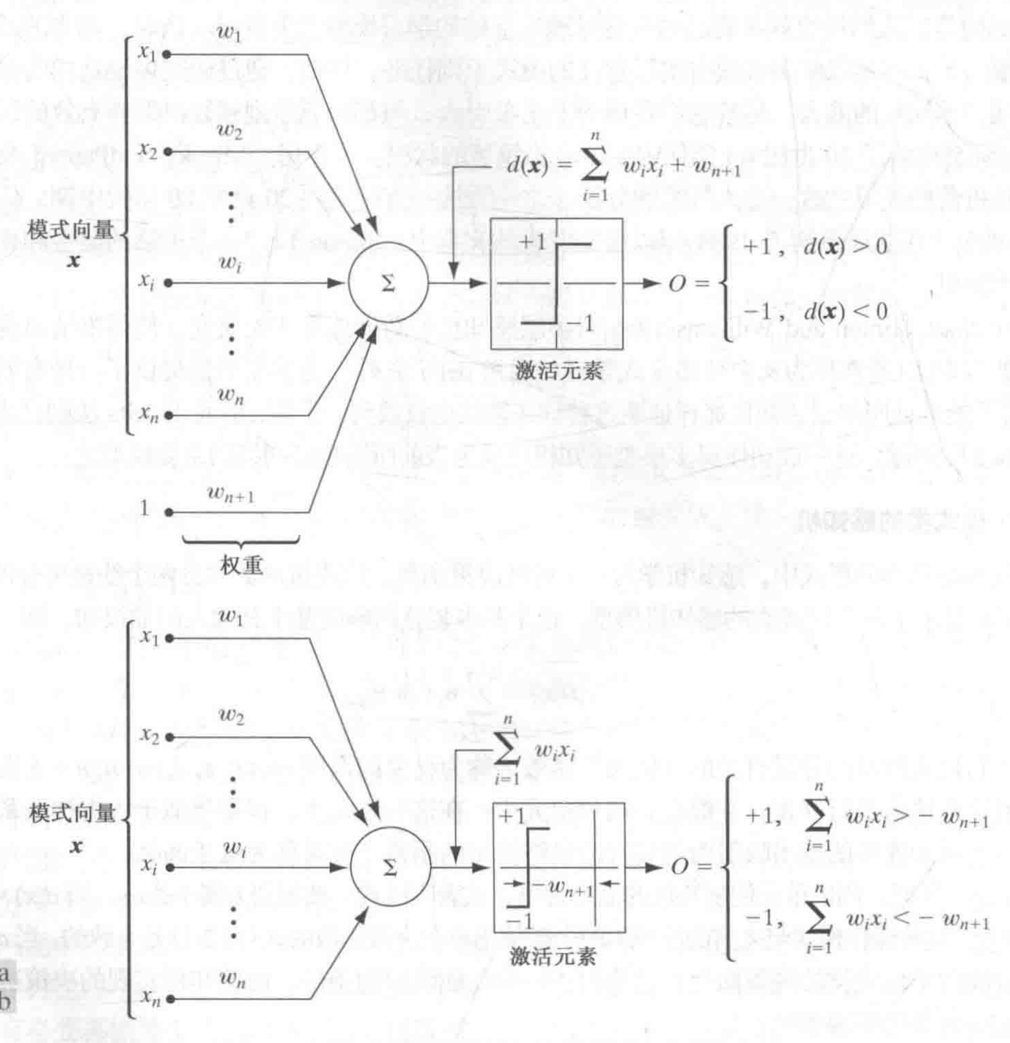


图 12.14　两个模式类的感知机模型的两种等效表示

另一个常用的公式是扩充这些模式向量，附加总等于 1 的额外的第$(n+1)$个元素，而不管类成员是什么。也就是说，扩充后的模式向量 $\boldsymbol{y}$ 是从模式向量 $\boldsymbol{x}$ 创建的，方法是令 $y_i=x_i,\ i=1,2,\cdots,n$ 并附加一个额外的元素 $y_{n+1}=1$。然后，式(12.2-29)变为

907

$$d(\boldsymbol{y})=\sum_{i=1}^{n+1} w_i y_i=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y} \tag{12.2-33}$$

其中 $\boldsymbol{y}=(y_1, y_2, \cdots, y_n, 1)^{\mathrm{T}}$ 现在是一个扩充的模式向量，而 $\boldsymbol{w}=(w_1, w_2, \cdots, w_n, w_{n+1})^{\mathrm{T}}$ 称为权向量。该表达式在表示时通常更为方便。然而，不管使用的公式是什么，关键问题都是使用来自两个类的一个给定模式向量训练集来找到 $\boldsymbol{w}$。

训练算法

下面的讨论中给出的算法，都是过去几年来为训练感知机提出的大量方法中的代表性方法。

线性可分的类：用于求两个线性可分训练集的权重向量解的一种简单迭代算法如下。对于两个分别属于类 ω_1 和类 ω_2 的扩充模式向量的训练集，令 $\boldsymbol{w}(1)$ 表示初始权重向量，它可能是任意选择的。然后，在第 k 步迭代中，如果 $\boldsymbol{y}(k)\in\omega_1$ 且 $\boldsymbol{w}^{\mathrm{T}}(k)\boldsymbol{y}(k)\leqslant 0$，则使用下式代替 $\boldsymbol{w}(k)$：

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+c\boldsymbol{y}(k) \tag{12.2-34}$$

式中，c 是一个正的修正增量。相反，如果 $\boldsymbol{y}(k)\in\omega_2$ 且 $\boldsymbol{w}^{\mathrm{T}}(k)\mathbf{y}(k)\geqslant 0$，则使用下式代替 $\boldsymbol{w}(k)$：

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)-c\boldsymbol{y}(k) \tag{12.2-35}$$

否则，保持 $\boldsymbol{w}(k)$ 不变：

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k) \tag{12.2-36}$$

该算法仅当正考虑模式在训练序列第 k 步被错误分类时，才改变 $\boldsymbol{w}$。假设修正增量 c 为正，现在它是一个常量。该算法有时称为*固定增量校正准则*。

当两个类的整个训练集循环通过机器而不出现任何错误时，该算法收敛。如果模式的两个训练集是线性可分的，那么固定增量校正准则会在有限步内收敛。称为*感知机训练定理*的该结果的证明，请参阅 Duda, Hart and Stork[2001]，Tou and Gonzalez[1974]和 Nilsson[1965]。

例 12.5　感知机算法示例。

考虑图 12.15(a) 中的两个训练集，每个训练集由两个模式组成。因为两个训练集都是线性可分的，所以该训练算法将会成功。在应用算法之前，模式先被扩充，对类 ω_1 生成训练集 $\{(0,0,1)^{\mathrm{T}},(0,1,1)^{\mathrm{T}}\}$，对类 ω_2 生成训练集 $\{(1,0,1)^{\mathrm{T}},(1,1,1)^{\mathrm{T}}\}$。令 $c=1$, $\boldsymbol{w}(1)=\mathbf{0}$，并按如下步骤顺序提交这些模式：

$$\boldsymbol{w}^{\mathrm{T}}(1)\boldsymbol{y}(1)=[0,0,0]\begin{bmatrix}0\\0\\1\end{bmatrix}=0 \qquad \boldsymbol{w}(2)=\boldsymbol{w}(1)+\boldsymbol{y}(1)=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\boldsymbol{w}^{\mathrm{T}}(2)\boldsymbol{y}(2)=[0,0,1]\begin{bmatrix}0\\1\\1\end{bmatrix}=1 \qquad \boldsymbol{w}(3)=\boldsymbol{w}(2)=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\boldsymbol{w}^{\mathrm{T}}(3)\boldsymbol{y}(3)=[0,0,1]\begin{bmatrix}1\\0\\1\end{bmatrix}=1 \qquad \boldsymbol{w}(4)=\boldsymbol{w}(3)-\boldsymbol{y}(3)=\begin{bmatrix}-1\\0\\0\end{bmatrix}$$

$$\boldsymbol{w}^{\mathrm{T}}(4)\boldsymbol{y}(4)=[-1,0,0]\begin{bmatrix}1\\1\\1\end{bmatrix}=-1 \qquad \boldsymbol{w}(5)=\boldsymbol{w}(4)\begin{bmatrix}-1\\0\\0\end{bmatrix}$$

其中，如式(12.2-34)和式(12.2-35)指出的那样，由于错误分类，在第一步和第三步中进行了权重向量修

907～909

正。由于仅当算法对所有训练模式完全无误迭代时，才能得到一个解，故必须再次提交训练集。机器学习过程按如下方式继续：令 $\boldsymbol{y}(5)=\boldsymbol{y}(1), \boldsymbol{y}(6)=\boldsymbol{y}(2), \boldsymbol{y}(7)=\boldsymbol{y}(3), \boldsymbol{y}(8)=\boldsymbol{y}(4)$，并以同样的方式进行。$k=14$ 时收敛，得到的权重向量解为 $\boldsymbol{w}(14)=(-2, 0, 1)^{\mathrm{T}}$。相应的决策函数是 $d(\boldsymbol{y})=-2y_1+1$。通过令 $x_i=y_i$ 返回原始模式空间，得到 $d(\boldsymbol{x})=-2x_1+1$，当集合等于 0 时，它变成图 12.15(b)所示的决策边界的方程。

不可分的类：实际上，线性可分的模式类是个例外，通常情况并非如此。因此，20 世纪 60 年代和 70 年代，大量研究工作转向处理不可分模式类的技术开发。随着神经网络训练方面取得进展，解决不可分行为的许多方法已成为仅具有历史意义的课题。然而，一种早期的方法与该讨论直接相关，这就是原始的 delta 规则。称为感知机训练的 Widrow-Hoff 或最小均方(LMS) delta 规则，在任何训练步骤都会使得实际响应与期望响应间的误差最小。

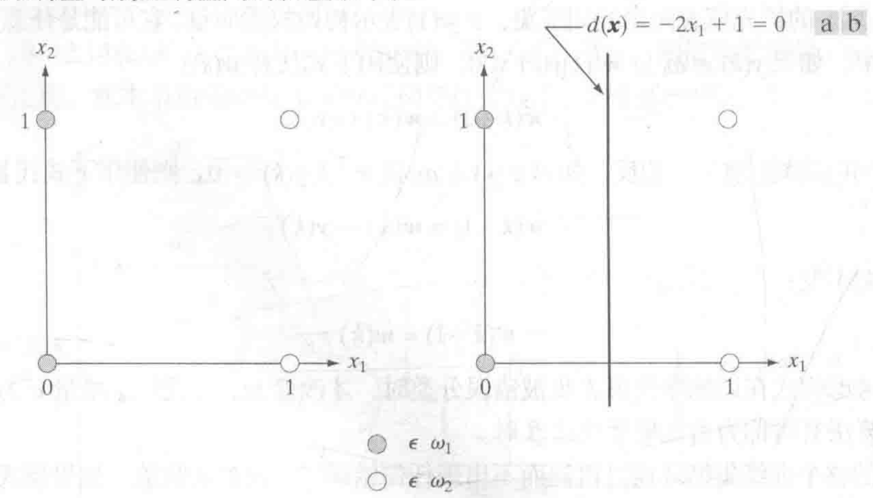


图 12.15 (a)属于两个类的模式；(b)由训练确定的决策边界

考虑准则函数

$$J(\boldsymbol{w})=\frac{1}{2}(r-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y})^2 \tag{12.2-37}$$

式中，r 是期望的响应(即当扩充后的训练模式向量 $\boldsymbol{y}$ 属于类 ω_1 时有 $r=+1$，而当 $\boldsymbol{y}$ 属于类 ω_2 时有 $r=-1$)。为在 $r=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}$ 时求该函数的最小值，其任务是在 $J(\boldsymbol{w})$的相反梯度方向逐步调整 $\boldsymbol{w}$，即该最小值对应于正确的分类。如果 $\boldsymbol{w}(k)$表示第 k 步迭代的权重向量，则通用梯度下降算法可写为

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)-\alpha\left[\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{w}}\right]_{\boldsymbol{w}=\boldsymbol{w}(k)} \tag{12.2-38}$$

其中 $\boldsymbol{w}(k+1)$是 $\boldsymbol{w}$ 的新值，而 $\alpha>0$ 给出修正量。由式(12.2-37)得到

$$\frac{\partial J(\boldsymbol{w})}{\partial \boldsymbol{w}}=-(r-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{y})\boldsymbol{y} \tag{12.2-39}$$

把这一结果代入式(12.2-38)得到

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\alpha\left[r(k)-\boldsymbol{w}^{\mathrm{T}}(k)\boldsymbol{y}(k)\right]\boldsymbol{y}(k) \tag{12.2-40}$$

其中初始权重向量 $\boldsymbol{w}(1)$是任意的。


909 ~ 910

通过将权重向量中的增量(delta)定义为

$$\Delta\boldsymbol{w}=\boldsymbol{w}(k+1)-\boldsymbol{w}(k) \tag{12.2-41}$$

我们可以按 delta 修正算法的形式将式(12.2-40)写为

$$\Delta \boldsymbol{w} = \alpha e(k)\boldsymbol{y}(k) \tag{12.2-42}$$

式中，

$$e(k) = r(k) - \boldsymbol{w}^{\mathrm{T}}(k)\boldsymbol{y}(k) \tag{12.2-43}$$

当模式 $\boldsymbol{y}(k)$存在时，它是权重向量 $\boldsymbol{w}(k)$提交的误差。

式(12.2-43)给出了权重向量 $\boldsymbol{w}(k)$的误差。如果把它更改为 $\boldsymbol{w}(k+1)$，但保持模式不变，则误差变为

$$e(k) = r(k) - \boldsymbol{w}^{\mathrm{T}}(k+1)\boldsymbol{y}(k) \tag{12.2-44}$$

故误差的变化是

$$\begin{aligned}\Delta e(k) &= \left[r(k) - \boldsymbol{w}^{\mathrm{T}}(k+1)\boldsymbol{y}(k)\right] - \left[r(k) - \boldsymbol{w}^{\mathrm{T}}(k)\boldsymbol{y}(k)\right] \\ &= -\left[\boldsymbol{w}^{\mathrm{T}}(k+1) - \boldsymbol{w}^{\mathrm{T}}(k)\right]\boldsymbol{y}(k) \\ &= -\Delta \boldsymbol{w}^{\mathrm{T}}\boldsymbol{y}(k)\end{aligned} \tag{12.2-45}$$

但由于 $\Delta \boldsymbol{w} = \alpha e(k)\boldsymbol{y}(k)$，所以

$$\begin{aligned}\Delta e &= -\alpha e(k)\boldsymbol{y}^{\mathrm{T}}(k)\boldsymbol{y}(k) \\ &= -\alpha e(k)\|\boldsymbol{y}(k)\|^2\end{aligned} \tag{12.2-46}$$

因此，权重的变化将误差减小了 $\alpha\|\boldsymbol{y}(k)\|^2$ 倍。下一个输入模式开始新的自适应循环，将下一个误差减小 $\alpha\|\boldsymbol{y}(k+1)\|^2$ 倍，以此类推。

α 的选择控制着稳定性和收敛速度(Widrow and Stearns[1985])。稳定性要求 $0<\alpha<2$。α 的实际范围是 $0.1<\alpha<1.0$。尽管这里未给出证明，但式(12.2-40)或式(12.2-42)和式(12.2-43)的算法确实收敛于一个解，这个解会使得在该训练集的模式上的均方误差最小。当模式类可分时，使用刚才讨论的算法得出的解，可能会也可能不会产生一个分离超平面。也就是说，在感知机训练理论的意义上，均方误差解并不意味着真有一个解。这种不确定性是使用这个特殊的公式在可分情形下和不可分情形下收敛算法的代价。

到目前为止，我们讨论的两个感知机训练算法可以推广到多个类及非线性决策函数。基于前面给出的历史评价，这里探讨多类训练算法并无多大价值。相反，我们将在神经网络的内容中讨论多类训练问题。

多层前馈神经网络

本节主要探讨多类模式识别问题的决策函数，并涉及由感知机计算单元的各个层组成的结构，与类是否可分无关。

基本结构：图12.16显示了所考虑神经网络模型的**结构**。它由多层结构上相同的计算节点(神经元)排列而成，从而一层中的每个神经元的输出送到下一层的每个神经元的输入。称为层 A 的第一层中的神经元的个数为 N_A。通常，$N_A = n$，它是输入模式向量的维度。称为 Q 层的输出层中的神经元的数量表示为 N_Q。N_Q 等于 W，即神经网络经训练后用于识别模式类的数量。如下面的讨论所示，如果该网络的第 i 个输出为“高”，而其他输出为“低”，则网络将模式向量 $\boldsymbol{x}$ 识别为属于类 ω_i。

如图 12.16 中的放大部分所示，每个神经元的形式都与前面讨论的感知机模式相同(见图 12.14)，只是硬性受限激活函数已被替代为软性受限“S 形”函数。在开发训练规则时，需要有沿神经网络所有路径的可微分性。下面的“S 形”激活函数具有必要的可微分性：

910 ~ 911

$$h_j(I_j) = \frac{1}{1 + \mathrm{e}^{-(I_j + \theta_j)/\theta_o}} \tag{12.2-47}$$

式中 $I_j, j = 1, 2, \cdots, N_J$ 是该网络第 J 层中的每个节点的激活元素的输入，θ_j 是偏移量，θ_o 控制“S 形”函数的形状。

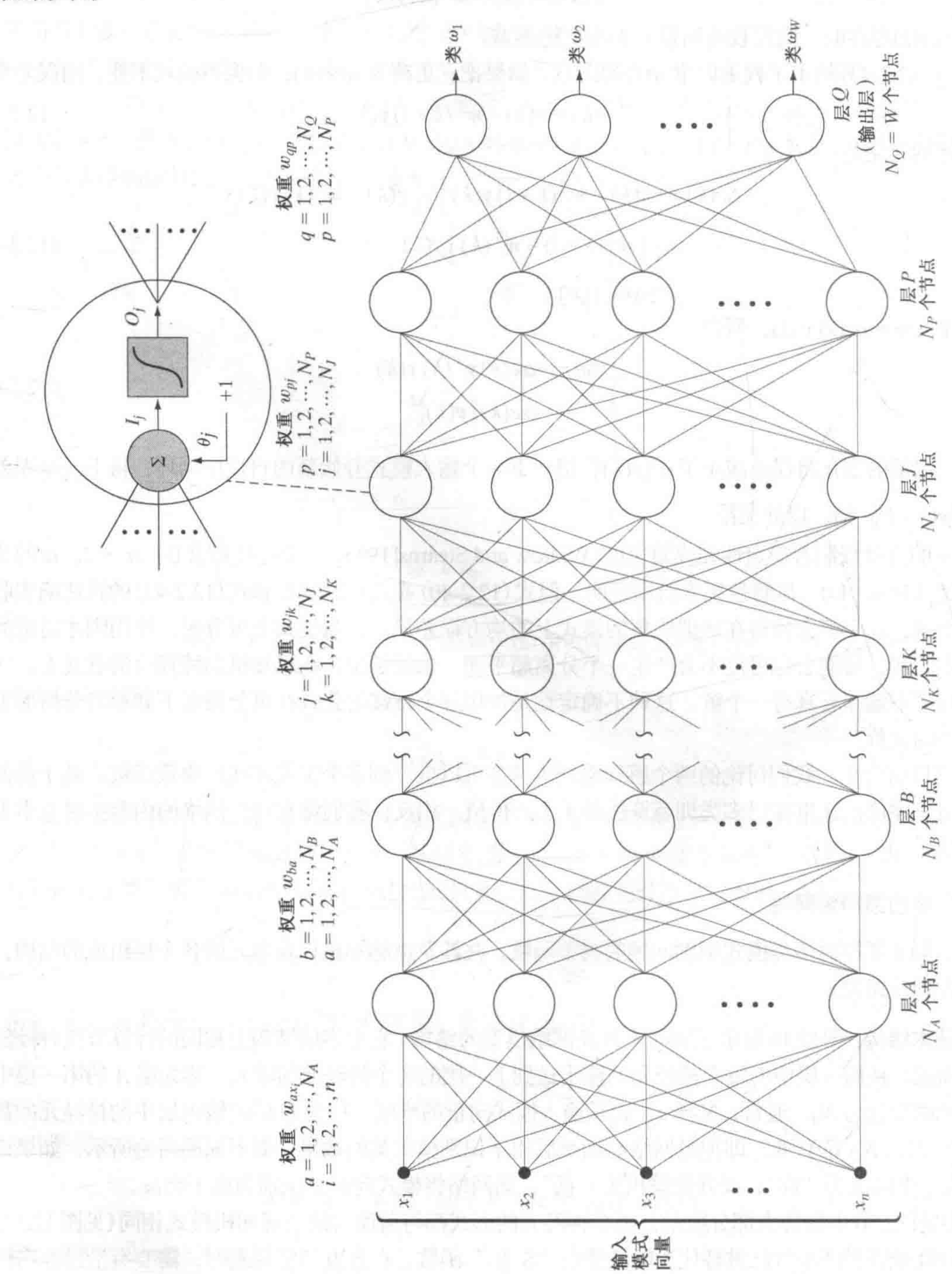


图 12.16 多层前馈神经网络模型。放大部分显示了整个网络中每个神经元的基本结构。偏移量 θ_j 仅当做另一个权重来处理

911 ~ 913

图 12.17 中画出了式(12.2-47)的曲线，并给出了每个节点的“高”响应限和“低”响应限。因此，在使用这个特殊的函数时，对所有大于 θ_j 的 I_j 值，系统会输出一个“高”读数。同样，对所有小于 θ_j 的 I_j 值，系统会输出一个“低”读数。如图12.17所示，“S 形”激活函数总为正，并且仅当激活元素的输入分别为负无穷或正无穷时，才到达其极限值 0 和 1。因此，在图 12.16 中，接近 0 和 1 的值(譬如 0.05 和 0.95)定义了神经元输出的“低”值和“高”值。原理上，不同类型的激活函数可用于不同的层，甚至用于神经网络的同一层中的不同节点。实际上，常用的方法是对整个网络使用相同形式的激活函数。

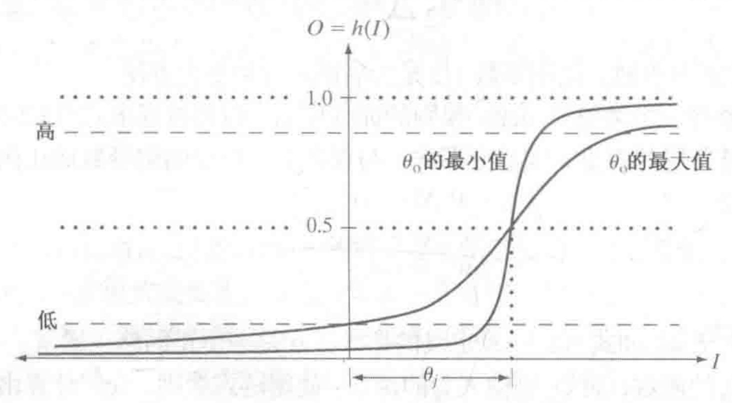


图 12.17　式(12.2-47)给出的“S 形”激活函数

参考图12.14(a)，图12.17 中显示的偏移量 θ_j 类似于前面讨论感知机时的权重系数 w_{n+1}。这个偏移阈值函数可采用图 12.14(a)的形式来实现，方法是将偏移 θ_j 作为一个额外的系数，然后使用该系数来修正统一输入到网络中所有节点的一个常量。为遵循文献中普遍使用的符号表示，我们未把单独的常量输入+1 显示在图12.16 中的所有节点上。取而代之，这个输入及其修正权重 θ_j 是网络节点的求和部分。如图 12.16 中放大部分显示的那样，在 J 层，N_J 个节点中的每个节点都有一个这样的系数。

在图 12.16 中，任何层中的一个节点的输入，都是来自前一层的输出的加权和。令层 K 表示层 J 的前一层(图 12.16 中未按字母顺序来划分层)，层 K 为层 J 中的每个节点的激活元素提供输入 I_j:

$$I_j = \sum_{k=1}^{N_K} w_{jk} O_k \tag{12.2-48}$$

式中 $j = 1, 2, \cdots, N_J$，这里 N_J 是层 J 中的节点数，N_K 是层 K 中的节点数，w_{jk} 是层 K 中的节点的输出 O_k 在送入层 J 中的节点之前，修正这些输出的权重。层 K 的输出为

$$O_k = h_k(I_k) \tag{12.2-49}$$

其中 $k = 1, 2, \cdots, N_K$。

清楚地理解式(12.2-48)中所用的下标符号很重要，因为我们在本节的剩余内容中都会用到它。首先，注意 $I_j, j = 1, 2, \cdots, N_J$ 表示层 J 中第 j 个节点的激活元素的输入。因此，I_1 表示层 J 中第 1 个(顶端)节点的激活元素的输入，I_2 表示层 J 中第 2 个节点的激活元素的输入，以此类推。层 J 中的每个节点都有 N_K 个输入，但对各个输入的加权是不同的。因此，层 J 中第 1 个节点的 N_K 个输入由系数 $w_{1k}, k = 1, 2, \cdots, N_K$ 加权；层 J 中第 2 个节点的输入由系数 $w_{2k}, k = 1, 2, \cdots, N_K$ 加权；以此类推。因此，层 K 的输出在送入层 J 时，需要指定共 $N_J \times N_K$ 个加权系数。完全指定层 J 中的节点，需要另外 N_J 个偏移系数 θ_j。

将式(12.2-48)代入式(12.2-47)得到

$$h_j(I_j) = \frac{1}{1 + \mathrm{e}^{-\left(\sum_{k=1}^{N_K} w_{jk} O_k + \theta_j\right) / \theta_o}} \tag{12.2-50}$$

这是本节剩余部分中要用到的激活函数的形式。

913 ~ 914

在训练过程中，修改输出层中的神经元很简单，因为每个节点的期望输出是已知的。训练一个多层网络的主要问题在于调整隐藏层中的权重，即那些非输出层的层中的权重。

反向传播训练： 我们先来关注输出层。(输出)层 Q 中各节点的期望响应 r_q 和相应真实响应 O_q 之间的总误差的平方为

$$E_Q = \frac{1}{2}\sum_{q=1}^{N_Q}(r_q - O_q)^2 \tag{12.2-51}$$

其中 N_Q 是输出层 Q 中的节点数，使用系数 1/2 是为后面取导数更为方便。

我们的目的是，推导一个类似于 delta 规则的训练规则，以便按寻求式(12.2-51)所示误差函数的最小值的方法，来调整每层的权重。像前面那样，与误差关于权重的偏导数成比例地调整权重，可实现这一目的。换句话说，

$$\Delta w_{qp} = -\alpha \frac{\partial E_Q}{\partial w_{qp}} \tag{12.2-52}$$

其中，层 P 先于层 Q，Δw_{qp} 如式(12.2-42)定义的那样，α 是一个正的修正增量。

误差 E_Q 是输出 O_q 的函数，而 O_q 是输入 I_q 的函数。使用链式法则，我们计算出 E_Q 的偏导数如下：

$$\frac{\partial E_Q}{\partial w_{qp}} = \frac{\partial E_Q}{\partial I_q}\frac{\partial E_q}{\partial w_{qp}} \tag{12.2-53}$$

由式(12.2-48)可得

$$\frac{\partial I_q}{\partial w_{qp}} = \frac{\partial}{\partial w_{qp}}\sum_{p=1}^{N_p} w_{qp}O_p = O_p \tag{12.2-54}$$

将式(12.2-53)和式(12.2-54)代入式(12.2-52)，得到

$$\Delta w_{qp} = -\alpha \frac{\partial E_Q}{\partial I_q} O_p = \alpha \delta_q O_p \tag{12.2-55}$$

式中，

$$\delta_q = -\frac{\partial E_Q}{\partial I_q} \tag{17.2-56}$$

为计算偏导数 $\partial E_Q / \partial I_q$，根据 E_Q 相对于 O_q 的变化率和 O_q 相对于 I_q 的变化率，我们使用链式法则表示偏导数，即

$$\delta_q = -\frac{\partial E_Q}{\partial I_q} = -\frac{\partial E_Q}{\partial O_q}\frac{\partial O_q}{\partial I_q} \tag{12.2-57}$$

由式(12.2-51)得到

$$\frac{\partial E_Q}{\partial O_q} = -(r_q - O_q) \tag{12.2-58}$$

而由式(12.2-49)得到

$$\frac{\partial O_q}{\partial I_q} = \frac{\partial}{\partial I_q}h_q(I_q) = h'_q(I_q) \tag{12.2-59}$$



914 ~ 916

将式(12.2-58)和式(12.2-59)代入式(12.2-57)得到

$$\delta_q = (r_q - O_q)h'_q(I_q) \tag{12.2-60}$$

它与误差值(r_q-O_q)成正比。将式(12.2-56)到式(12.2-58)代入式(12.2-55)，最终得到

$$\Delta w_{qp}=\alpha(r_q-O_q)h'_q(I_q)O_p=\alpha\delta_q O_p \tag{12.2-61}$$

指定了函数$h_q(I_q)$后，式(12.2-61)中的所有项都是已知的，或者可以在网络中观察到。换句话说，根据对网络输入的任何训练模式的表示，我们就知道每个输出节点的期望响应r_q应是什么。就像层Q的激活元素的输入I_q和层P中节点的输出O_p那样，每个输出节点的值O_q是可观察到的。这样，我们就知道如何调整权重，从而改进网络中最后一层和上一层之间的链接。

将输出层放到一边，我们现在分析层P中发生的情况。采用与上面相同的方式，可得

$$\Delta w_{pj}=\alpha(r_p-O_p)h'_p(I_p)O_j=\alpha\delta_p O_j \tag{12.2-62}$$

其中误差项是

$$\delta_p=(r_p-O_p)h'_p(I_p) \tag{12.2-63}$$

除了r_p外，式(12.2-62)和式(12.2-63)中的所有项要么是已知的，要么可在网络中观察到。在内部层中r_p项没有意义，因为根据模式成员我们确实不知道一个内部节点的响应是什么。仅在模式分类最终发生的网络的输出处，才能指定想要的响应r。如果知道内部节点的信息，就不需要更多的层。因此，我们必须根据已知的或可在网络中观察到的量来找到重新定义δ_p的一种方法。

回到式(12.2-57)，我们将层P的误差项写为

$$\delta_p=-\frac{\partial E_p}{\partial I_p}=\frac{\partial E_p}{\partial O_p}\frac{\partial O_p}{\partial E_p} \tag{12.2-64}$$

项$\partial O_p/\partial I_p$不难表示，如先前那样，它是

$$\frac{\partial O_p}{\partial I_p}=\frac{\partial h_p(I_p)}{\partial I_p}=h'_p(I_p) \tag{12.2-65}$$

一旦确定了h_p，上式就是已知的，因为I_p可通过观察得到。产生的r_p项是偏导数$\partial E_p/\partial O_p$，所以该项必须以不包含$r_p$的方式来表示。使用链式法则，我们将偏导数写为

$$-\frac{\partial E_P}{\partial O_p}=-\sum_{q=1}^{N_Q}\frac{\partial E_P}{\partial I_q}\frac{\partial I_q}{\partial O_p}=-\sum_{q=1}^{N_Q}\left(\frac{\partial E_P}{\partial I_q}\right)\frac{\partial}{\partial O_p}\sum_{p=1}^{N_P}w_{qp}O_p=\sum_{q=1}^{N_Q}\left(-\frac{\partial E_P}{\partial I_q}\right)w_{qp}=\sum_{q=1}^{N_Q}\delta_q w_{qp} \tag{12.2-66}$$

其中最后一步来自式(12.2-56)。将式(12.2-65)和式(12.2-66)代入式(12.2-64)，得到期望的δ_p表达式为

$$\delta_p=h'_p(I_p)\sum_{q=1}^{N_Q}\delta_q w_{qp} \tag{12.2-67}$$

现在可以计算出参量δ_p，因为其所有项都是已知的。这样，式(12.2-62)和式(12.2-67)就完全为层P建立了训练规则。式(12.2-67)的重要性是由量δ_q和w_{qp}来计算δ_p，而这两个量是在紧邻层P的层中计算出来的项。计算出层P的误差项和权重后，这些量可类似地用于计算紧邻层P之前的层的误差项和权重。换句话说，我们找到了一种从输出层的误差开始，将误差反向传播回网络的方法。

我们可将训练过程总结并归纳如下。对于任何层K和J，其中层K是紧邻层J的前一层，计算权重w_{jk}，该权重使用下式修改这两层之间的连接：

$$\Delta w_{jk}=\alpha\delta_j O_k \tag{12.2-68}$$

如果层J是输出层，则δ_j为

$$\delta_j=(r_j-O_j)h'_j(I_j) \tag{12.2-69}$$

916 ~ 917

如果层J是一个内部层，且层P是它的下一层(右侧)，则δ_j为

$$\delta_j = h_j'(I_j)\sum_{p=1}^{N_p}\delta_p w_{jp} \tag{12.2-70}$$

其中$j=1,2,\cdots,N_J$。使用式(12.2-50)中的激活函数并令$\theta_o=1$，可得

$$h_j'(I_j) = O_j(1-O_j) \tag{12.2-71}$$

在这种情况下，式(12.2-69)和式(12.2-70)采用特别有吸引力的形式，即对于输出层有

$$\delta_j = (r_j - O_j)O_j(1-O_j) \tag{12.2-72}$$

而对于内部层有

$$\delta_j = O_j(1-O_j)\sum_{p=1}^{N_p}\delta_p w_{jp} \tag{12.2-73}$$

在式(12.2-72)和式(12.2-73)中，都有$j=1,2,\cdots,N_J$。

式(12.2-68)到式(12.2-70)组成了图12.16中多层前馈神经网络的一般delta训练法则。该处理由整个网络的任意一组(都不相等的)权重开始。然后，在任意迭代步骤中应用这个一般delta法则包括两个基本阶段。在第一个阶段，为网络提供一个训练向量，通过允许该向量在各个层中传播来计算出每个节点的输出Q_j。然后，将输出层中各个节点的输出O_q与它们的期望响应r_p进行比较，生成误差项δ_q。第二阶段包括一条通过网络的反向路径，在此期间，合适的误差信号被传递给每个节点，并做相应的权重变化。这一过程同样适用于偏置权重θ_j。如前面详细讨论的那样，这些权重被简单地当作额外的权重来处理，它们把一个单元输入修改为网络中每个节点的求和连接。

通常的做法是，追踪该网络误差以及与各个模式相关联的误差。在成功的训练会话中，网络误差会随迭代次数的增加而减小，并且该过程会收敛到一个稳定的权重集合，对于附加的训练，这些权重仅表现为很小的波动。紧随着在训练期间完成的一个模式是否被正确分类的方法是，从得到的模式确定输出层的响应，如以前定义的那样，与模式类有关的输出层节点的响应为高，所有其他节点的输出为低。

系统被训练之后，就会使用训练阶段中建立的参数来对模式进行分类。在正常操作中，所有的反馈路径是被分开的。然后，允许任何输入模式传过不同的层，且该模式被分类为属于具有高值输出节点的类，此时，其他所有节点输出为低。如果被标记为高的节点不止一个，或没有节点输出被标记为高，则将其声明为一个错误分类，或简单地将该模式赋给具有最高数值的输出节点的类。

例12.6　使用神经网络的形状分类。

我们现在说明如何训练一个如图12.16所示的神经网络，以便识别图12.18(a)中所示的4种形状，以及图12.18(b)中所示的这4种形状的带有噪声形式的样本。

通过计算这些形状的归一化信号(见11.1.3节)产生模式向量，然后得到每个信号的48个均匀间隔的样本。将得到的48维向量作为图12.19所示的三层前馈神经网络的输入。第一层中神经元节点数选择为48，它对应于输入模式向量的维数。第三层(输出层)中的4个神经元对应于模式类的数量，而中间层的神经元的数量指定为26(输入层和输出层中的神经元的平均数)。由于不知道确定神经网络内部层的节点数的规则，因此该数量通常要么基于先验知识，要么简单地任意选择，然后通过测试来完善。在输出层中，从上到下的4个节点此时分别代表类$\omega_j, j=1,2,3,4$。在设定网络结构后，必须为每个单元和层选定激活函数。根据前面的讨论，所有选定的激活函数都要满足式(12.2-50)，其中$\theta_o=1$，以便可以应用式(12.2-72)和式(12.2-73)。

917~919

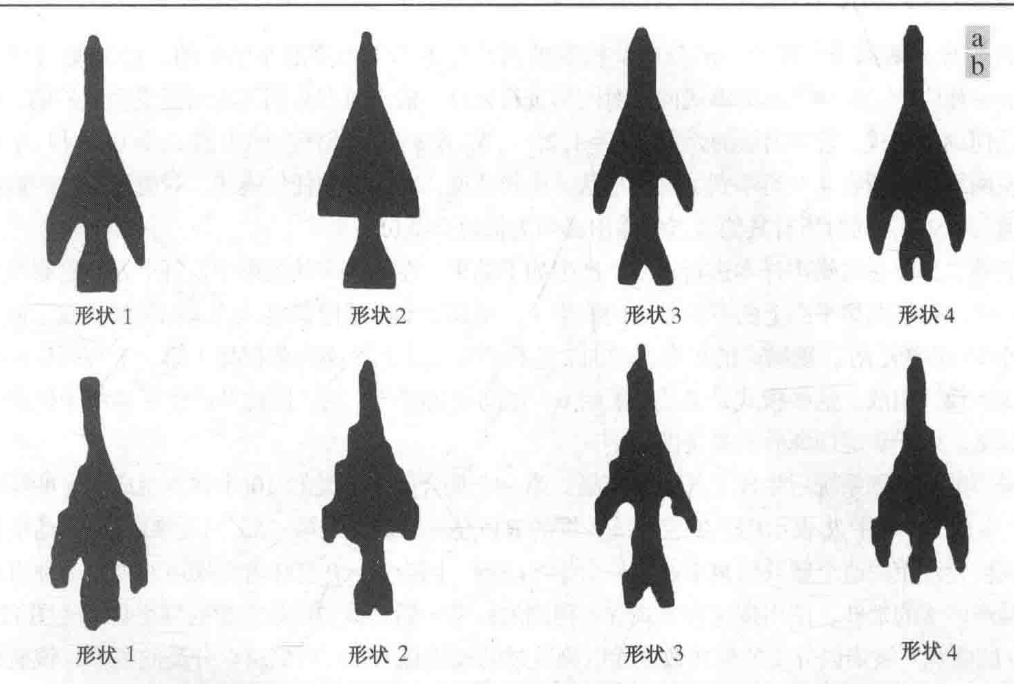


图 12.18 训练图 12.19 中的神经网络时使用的(a)参考图形和(b)典型带噪声图形(图像由南伊利诺依大学 ECE 系的 Lalit Gupta 博士提供)

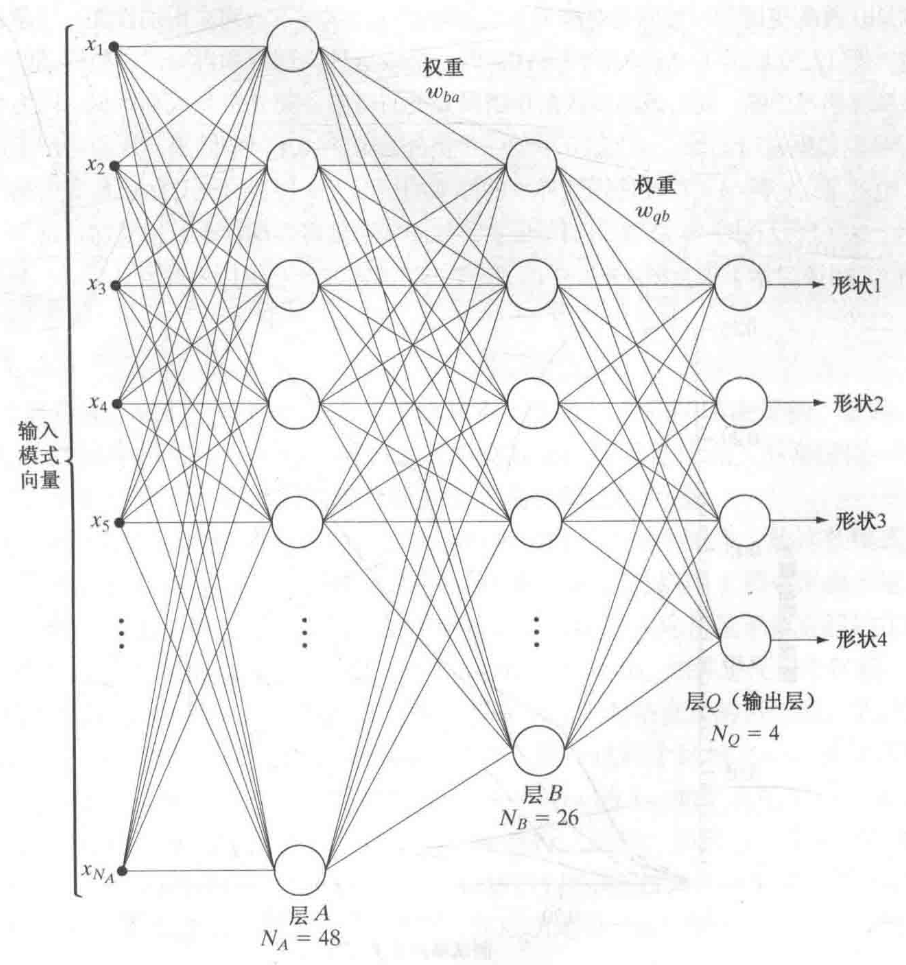


图 12.19 用于识别图 12.18 中所示形状的三层神经网络(图像由南伊利诺伊大学 ECE 系的 Lalit Gupta 博士提供)

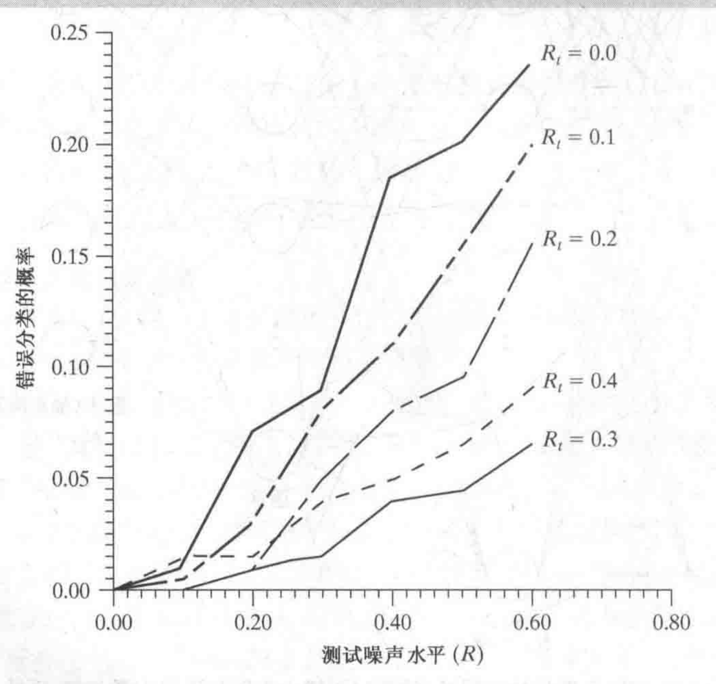
919
~
920

训练过程分为两部分。在第一部分中，权重被初始化为均值为零的小随机值，然后使用与类似于图 12.18(a)中相应的无噪声样本的模式向量对网络进行训练。输出节点在训练期间是受到监控的。对于来自类ω_i的任何训练模式，若输出层的元素在$q=1,2,\cdots,N_Q$且$q\neq i$时，产生输出值$O_i\geqslant 0.95$和$O_q\leqslant 0.05$，那么我们说网络已从所有 4 个类学到了这些形状。换句话说，对类ω_i的任何模式，对应于该类的输出单元必须为高值(≥0.95)，同时所有其他节点的输出必须为低值(≤0.05)。

训练的第二部分是对噪声样本进行训练，产生如下结果。在无噪声的图形中，每个处在轮廓线上的像素被赋予一个保留在图像平面上的原始坐标的概率V，而其 8 个邻域像素之一的坐标被随机赋予概率$R=1-V$。减小V(即增大R)，则噪声的程度会增加。这样就生成了两个噪声数据集。第一个噪声集由每个类的 100 个噪声模式组成，这些模式是R在 0.1 到 0.6 之间变化时产生的，因此共产生了 400 个模式。该集合称为测试集，用于确定训练后的系统性能。

使用噪声数据训练系统后生成了几个噪声集。第一个集合由每个类的 10 个样本组成，这些样本是通过令$R_t=0$生成的，其中R_t表示用于生成训练数据的R的值。从训练的第一部分(无噪声)得到的权重向量开始，系统被允许使用这个新数据集来遍历一个学习序列。因为$R_t=0$意味着无噪声，因此这种再训练是之前的无噪声训练的延伸。使用按这种方式学习得到的权重，满足该测试数据集的网络会产生图 12.20 中标有$R_t=0$的曲线。被错误分类的模式数量除以测试过的模式总数，可得到错误分类的概率，该概率通常是用于建立神经网络性能的一个测度。

接下来，从使用$R_t=0$生成的数据学习到的权重向量开始，使用由$R_t=0.1$生成的噪声数据集合对系统重新进行训练。再次使用新权重向量对整个系统运行这些测试样本，建立识别性能。注意，系统在性能上的显著改进。图 12.20 显示了$R_t=0.2, 0.3$和 0.4 时，继续这种再训练和再测试过程得到的结果。如期望的那样，如果系统学习正确，则由该测试数据集错误分类的概率会随R_t的增大而降低，因为系统使用了具有较高R_t值的噪声数据进行训练。图 12.20 中的一个例外是$R_t=0.4$时的结果。原因是用于训练该系统的样本数较少。也就是说，网络无法在使用这样少的样本的情况下，使自身充分适应具有更高噪声水平的形状的较大变化。这一假设在图 12.21 中得到了证实，该图表明随着训练样本数的增加，错误分类的概率会随着降低。图 12.21 还显示了来自图 12.20 中的$R_t=0.3$时的曲线，它仅作为参考。



919~921

图 12.20　神经网络的性能是噪声水平的函数(原图由南伊利诺伊大学 ECE 系的 Lalit Gupta 博士提供)

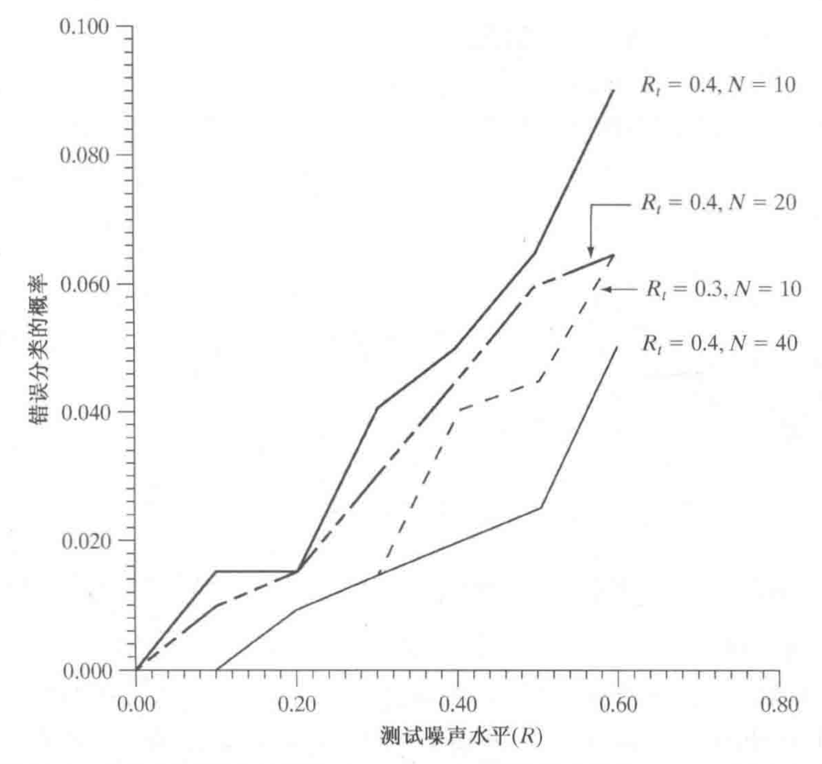


图 12.21　增加训练模式的数量时，$R_t = 0.4$ 时的性能改进($R_t = 0.3$ 时的曲线仅作为参考)(原图由南伊利诺伊大学 ECE 系的 Lalit Gupta 博士提供)

前述结果表明，三层神经网络经过适当水平的训练后，具有学习识别被噪声污染的形状能力。甚至当使用无噪声数据(图 12.20 中 $R_t = 0$)训练时，使用被噪声高度污染的数据(图 12.20 中 $R = 0.6$)进行测试，系统也能实现接近 77%的正确识别水平。当系统使用噪声水平更高的数据($R_r = 0.3$ 和 $R = 0.4$)训练后，对相同数据的识别率增加到 99%。注意，系统地、小量地增大附加噪声来训练系统，可增大系统的分类能力，这一点很重要。已知噪声的性质时，这种方法对改进神经网络在学习过程中的收敛性和稳定性是很理想的。

决策面的复杂性：我们已经证明一个单层感知机可实现一个超平面决策面。此时，一个很自然的问题是：由图12.16中模型的一个多层网络实现的决策面的性质是什么？下面讨论一个三层网络能实现由相交的超平面组成的任意复杂的决策面，由此说明前述的问题。

首先，考虑图 12.22(a)中的一个双输入的双层网络。由于有两个输入，故这些模式是二维的，从而网络的第一层中的每个节点在二维空间中实现一条直线。我们分别用 1 和 0 来表示这两个节点的高输出和低输出。假设输出 1 指出对应于第一层中的一个节点的输入向量位于该直线的正侧。那么馈送到第二层中的单个节点的可能输出组合是(1, 1), (1, 0), (0, 1)和(0, 0)。如果定义两个区域，一个区域对应于类ω_1，位于两条直线的正侧；另一个区域对应于类ω_2，位于两条直线的另一侧，通过简单地执行逻辑“与”操作，输出节点就可以把任何输入模式分类为属于这两个区域之一。换句话说，仅当来自第一层的两个输出都为 1 时，输出节点的响应才为表示类ω_1的 1。如果 θ_j 的值位于半开区间(1, 2]，那么“与”操作可以由先前讨论的神经节点的形式来执行。这样，如果假设第一层的响应为 0 和 1，那么只有在来自第一层的该神经节点上执行的求和大于 1 时，输出节点的响应才是表示类ω_1的高值。图 12.22(b)和(c)显示了图12.22(a)所示的网络是如何成功地等分这两个模式类的，而使用单个线性面则做不到这一点。

921 ~ 923

如果第一层中的节点数增加到三个，那么图12.22(a)中的网络将可实现由三条直线相交组成的一个决策边界。类ω_1位于所有三条直线的正侧这一要求，将生成由这三条直线围成的一个凸形区域。事实上，任何开放的或闭合的凸区域，都可以简单地通过增加两层神经网络的第一层的节点数来构造。

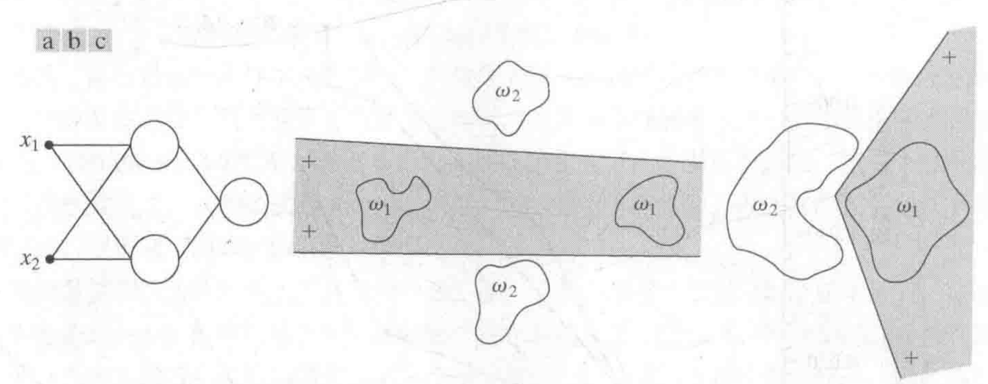


图 12.22 (a)一个双输入双层前馈神经网络；(b)和(c)可使用该网络实现的决策边界示例

下一个逻辑步骤是将层数增加到三层。像之前那样，此时第一层中的节点实现一条直线。为形成由各条直线围成的区域，第二层的节点执行“与”操作。第三层的节点将不同的类划归不同区域。例如，假设ω_1由两个不同的区域组成，每个区域由不同的一组直线围成边界，那么第二层中的两个节点用于对应于同一模式类的区域。当第二层中的两个节点之一变为高值时，输出节点之一应该能够提示该类的存在。假设第二层中的高和低状态分别表示为 1 和 0，那么让网络的输出节点执行逻辑“或”操作，可得到这种能力。根据前面讨论的神经节点的形式，我们可通过将θ_j的值设置在半开区间[0, 1)内来这样做。然后，只要第二层中至少有一个节点与变为高值(输出为 1)的输出节点相关联，输出层中的相应节点也将变为高值，从而将该模式分配给与该节点相关联的类。

图12.23总结了前面的讨论。注意，在第三行中，由三层网络实现的决策区域的复杂性在原理上是任意的。实际上，在构造可正确响应与特殊类相关联的各种组合的第二层时，通常会很困难。原因是直线不会刚好截止于与其他直线的交点处，从而同一类的模式可能会出现在该模式空间中的直线的两侧。实际上，对于一个给定的模式类，第二层可能很难断定哪些直线应该包含在“与”操作中，或许根本上就不可能。参考图12.23的第三列中的“异或”问题表明，如果输入模式是二元形式的，那么在两个维度的情形下只能构建 4 个不同的模式。如果模式按如下方式排列：类ω_1由模式{(0, 1), (1, 0)}组成，类ω_2由模式{(0, 0), (1, 1)}组成，那么这两个类中模式类的归属由“异或”(XOR)逻辑函数给出，仅当两个变量中的一个或另一个为 1 时，该函数值才为 1，而其他情形下它为 0。因此，XOR 函数值为 1 指出模式属于类ω_1，而 XOR 函数值为 0 指出模式属于ω_2。

前面的讨论可以直接推广到 n 维情形，此时我们使用超平面来替代直线。单层网络实现单个超平面。两层网络实现由超平面的交集组成的任意凸形区域。三层网络实现任意复杂度的决策面。每层中使用的节点数决定了后两种情形的复杂度。在第一种情形下，类的数量限制为两个。在其他两种情形下，类的数量是任意的，因为可以根据手边的问题来选择输出节点的数量。

考虑前述说明，一个合乎逻辑的问题是，为什么人们对研究三层以上的神经网络感兴趣？毕竟三层网络就可以实现任意复杂度的决策面。答案在于仅使用三层来训练网络的方法。用于图12.16中网络的训练规则可使误差最小，但未说明如何将一组超平面与之前讨论的三层网络的第二层中的节点关联起来。事实上，如何在层次数和每层中的节点数之间进行折中分析这一问题，仍有待解决。在实际中，通常通过试错法或给定的问题域的已有经验来解决分析中的问题。

923 ~ 924

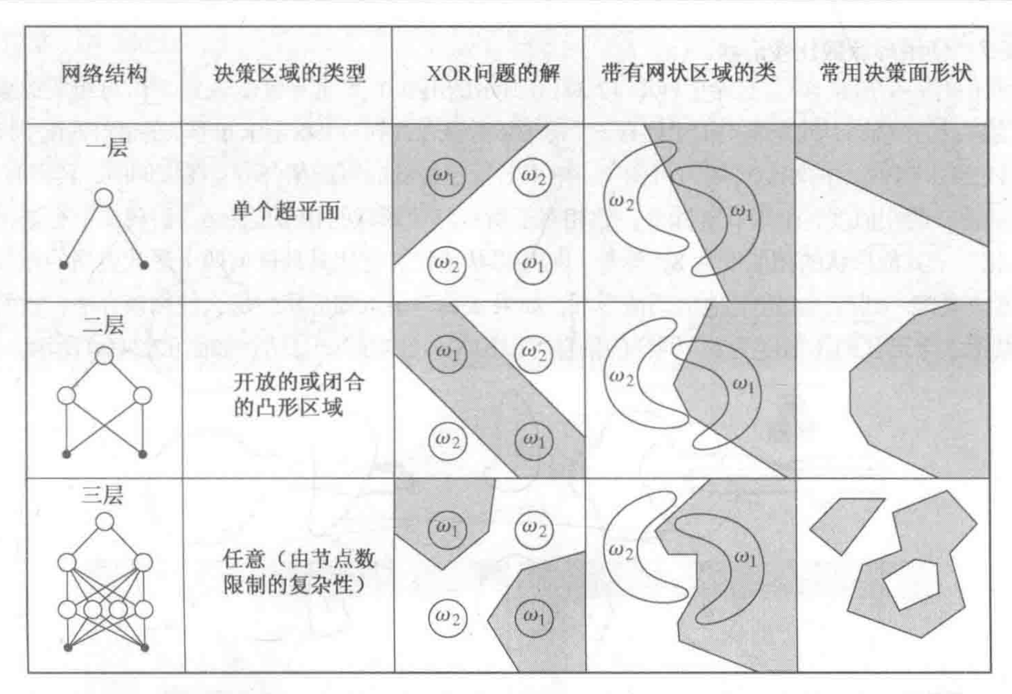


图 12.23　可以由单层和多层前馈网络带有一层或两层隐藏单元与两个输入形成的决策区域的类型(原图由 Lippman 提供)

12.3　结构方法

12.2 节中讨论的技术可定量地处理模式，并且大部分忽略了模式形状中固有的结构关系。然而，本节中讨论的结构方法可通过精确地运用这些类型的关系来实现模式识别的目的。本节介绍两种基本的基于串表示的边界形状的识别方法，串是结构模式识别中最实用的方法。

12.3.1　匹配形状数

为了比较根据形状数描述的区域边界，我们可以明确表达一个过程，该过程类似于 12.2.1 节中为模式向量引入的最小距离概念。参考 11.2.2 节中的讨论可知，两个区域边界(形状)之间的相似度 k 定义为它们的形状数仍保持一致的最大阶。例如，令 a 和 b 代表由 4 方向链码表示的闭合边界的形状数。如果

$$\begin{aligned} s_j(a) &= s_j(b), \qquad j = 4, 6, 8, \cdots, k \\ s_j(a) &\neq s_j(b), \qquad j = k+2, k+4, \cdots \end{aligned} \tag{12.3-1}$$

那么这两个形状有一个相似度 k，其中 s 代表形状数，下标代表阶。两个形状 a 和 b 之间的距离定义为它们的相似度的倒数：

$$D(a,b) = \frac{1}{k} \tag{12.3-2}$$

该距离满足如下性质：

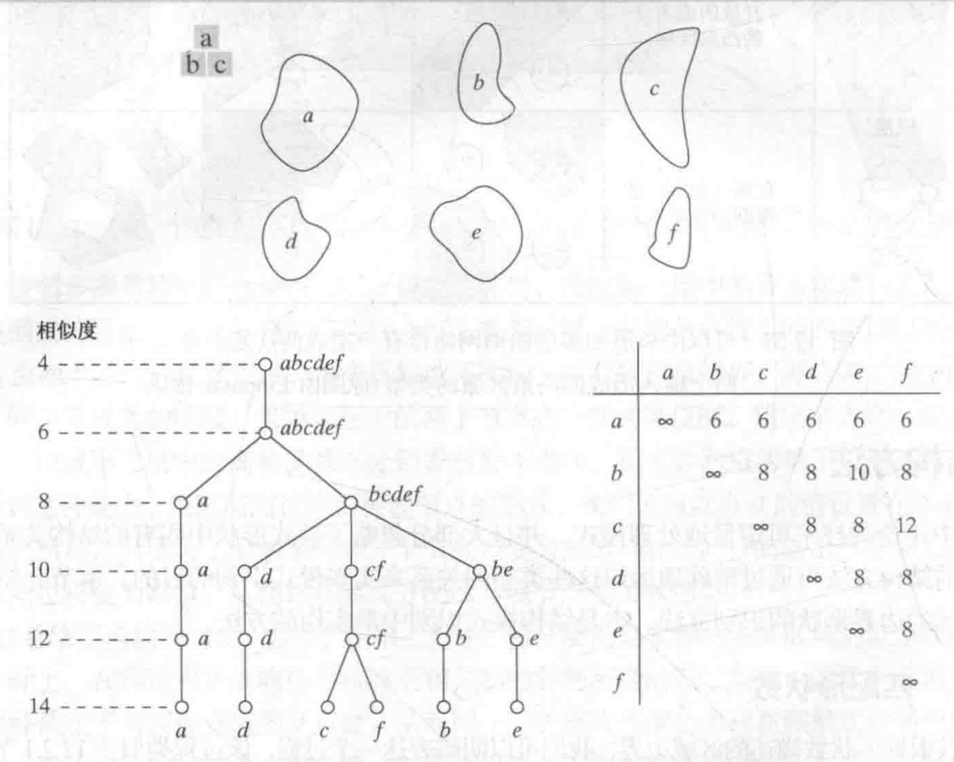
$$\begin{aligned} &D(a,b) \geqslant 0 \\ &D(a,b) = 0, \text{ iff } a = b \\ &D(a,c) \leqslant \max\left[(D(a,b), D(b,c)\right] \end{aligned} \tag{12.3-3}$$

k 或 D 都可以用于比较这两个形状。如果使用相似度，则 k 越大，形状就越相似(注意，对于相同的形状，k 为无穷大)。使用距离测度时，情形正好相反。

924
~
925

例 12.7 使用形状数比较形状。

假设我们有一个形状 f，并且希望在图 12.24(a)所示的由 5 个其他形状(a, b, c, d 和 e)组成的集合中找到与它最为接近的匹配。这个问题类似于拥有 5 个原形状并试图找到一个给定未知形状的最好匹配。该搜索可借助于图12.24(b)中所示的相似性树来可视化。树根对应于最低的可能相似度，在该例中，其为4。假设这些形状有高达 8 的相似度，但形状 a 除外，它相对于所有其他形状的相似度为 6。沿树向下行进，我们发现形状 d 相对于其他形状的相似度为 8，等等。如果形状 f 和 c 有比其他任何两个形状更高的相似度，那么形状 f 和 c 是唯一匹配的。在其他极端情形下，如果 a 是一个未知形状，那么使用该方法，我们可以说 a 类似于其他 5 个形状的相似度为 6。同样的信息可以用相似性矩阵来总结，如图 12.24(c)所示。



	a	b	c	d	e	f
a	∞	6	6	6	6	6
b		∞	8	8	10	8
c			∞	8	8	12
d				∞	8	8
e					∞	8
f						∞

图 12.24 (a)各种形状；(b)假想的相似树；(c)相似性矩阵(原图由 Bribiesca and Guzman 提供)

12.3.2 串匹配

假设两个区域边界 a 和 b 已被编码成串(见 11.5 节)，两个串分别表示为 $a_1, a_2, \cdots, a_n$ 和 $b_1, b_2, \cdots, b_m$。令α表示两个串之间的匹配数，如果 $a_k = b_k$，则匹配出现在第 k 个位置。不匹配的符号数为

$$\beta = \max\left(|a|,|b|\right) - \alpha \tag{12.3-4}$$

式中，|arg|是该宗量的串表示的长度(符号数)。可以证明，当且仅当 a 和 b 相同时 $\beta = 0$(见习题 12.21)。

a 和 b 间一种简单的相似性测度是比率

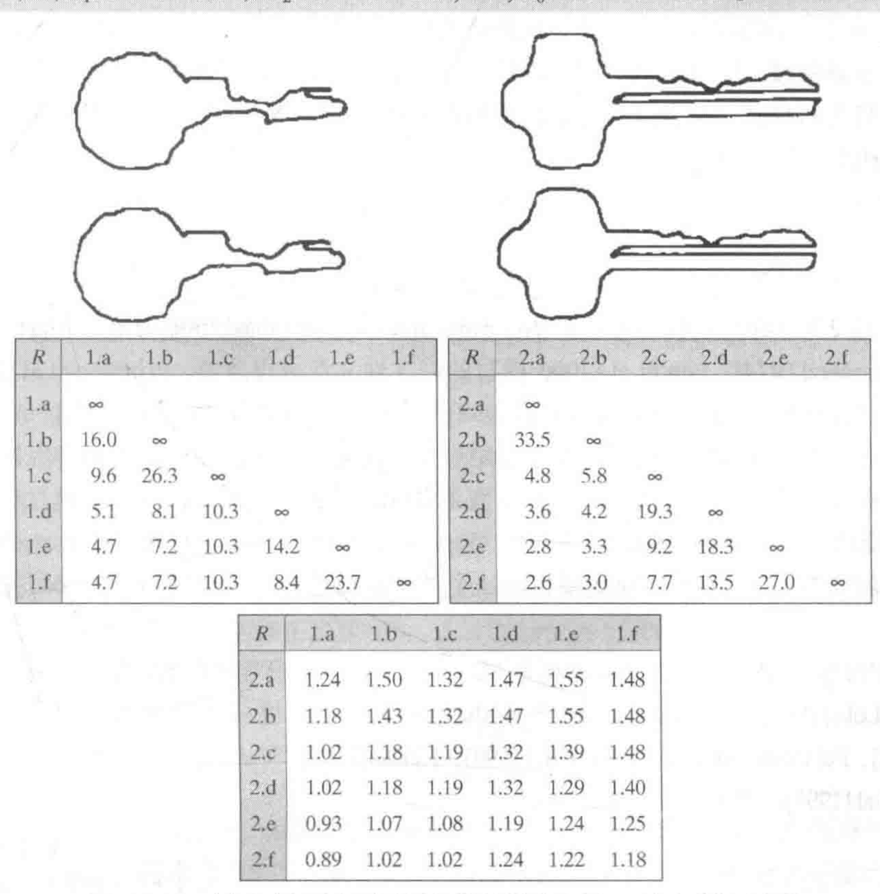
$$R = \frac{\alpha}{\beta} = \frac{\alpha}{\max\left(|a|,|b|\right) - \alpha} \tag{12.3-5}$$

对于完美匹配，R 为无限大，而当 a 和 b 中没有任何符号匹配时，R 为 0(此时 $\alpha = 0$)。因为匹配是逐个字符进行的，对于降低计算量来说，每条边界上的起始点就很重要。任何归一化到或近似归一化到相同起点的方法都有帮助，只要它与强力匹配相比可提供计算上的优势，强力搜索从每个串上的任意点开始，然后移位一个字符(采用卷绕的方式)并为每个移位计算式(12.3-5)。R 的最大值给出最好的匹配。

905~927

例 12.8　串匹配示例。

图 12.25(a)和(b)显示了来自两个目标类的样本边界，它们由一个多边形拟合来近似(见 11.1.3 节)。图 12.25(c)和(d)分别显示了对应于图12.25(a)和(b)所示边界的多边形近似。当顺时针追踪每个多边形时，通过计算各线段间的内角θ，就形成了来自这些多边形的串。这些角度被编码成 8 个可能的符号，角度增量为 45°，即α_1: $0° < \theta \leqslant 45°$，$\alpha_2$: $45° < \theta \leqslant 90°$，…，$\alpha_8$: $315° < \theta \leqslant 360°$。



R	1.a	1.b	1.c	1.d	1.e	1.f
1.a	∞					
1.b	16.0	∞				
1.c	9.6	26.3	∞			
1.d	5.1	8.1	10.3	∞		
1.e	4.7	7.2	10.3	14.2	∞	
1.f	4.7	7.2	10.3	8.4	23.7	∞

R	2.a	2.b	2.c	2.d	2.e	2.f
2.a	∞					
2.b	33.5	∞				
2.c	4.8	5.8	∞			
2.d	3.6	4.2	19.3	∞		
2.e	2.8	3.3	9.2	18.3	∞	
2.f	2.6	3.0	7.7	13.5	27.0	∞

R	1.a	1.b	1.c	1.d	1.e	1.f
2.a	1.24	1.50	1.32	1.47	1.55	1.48
2.b	1.18	1.43	1.32	1.47	1.55	1.48
2.c	1.02	1.18	1.19	1.32	1.39	1.48
2.d	1.02	1.18	1.19	1.32	1.29	1.40
2.e	0.93	1.07	1.08	1.19	1.24	1.25
2.f	0.89	1.02	1.02	1.24	1.22	1.18

图 12.25　(a)和(b)两个不同目标类的样本边界；(c)和(d)对应的多边形近似；(e)~(g)*R* 值表(原图由 Sze 和 Yang 提供)

图 12.25(e)显示了针对目标 1 本身的 6 个样本来计算测度 *R* 的结果。例如，对应于各个 *R* 值，符号 1.c 指的是目标类 1 的第三个串。图 12.25(f)显示了第二个目标类的串和它们自身进行比较的结果。最后，图 12.25(g)显示了通过比较一个类的串与其他类的串得到的 *R* 值表。注意，这里所有的 *R* 值比前两个表中的任何项都小，这表明 *R* 测度在两个目标类间的辨识达到了很高的程度。例如，如果串 1.a 的类别成员未知，那么通过将该串和类 1 的样本(原型)串相比较产生的最小 *R* 值将是 4.7 [见图 12.25(e)]。相比之下，将该串和类别 2 的串相比较产生的最大值将是 1.24 [见图 12.25(g)]。由此可得出如下结论：串 1.a 是目标类 1 的成员。这种分类方法类似于 12.2.1 节中介绍的最小距离分类器。

小结

从第 9 章开始，我们对数字图像处理的论述发生了改变，即从其输出为图像的处理变为 1.1 节中详细说明的其输出为图像属性的处理。尽管本章中的内容本质上是介绍性的，但所覆盖的主题对于理解目标识别技术相当基础。如本章开头提及的那样，本章以目标的识别来结束合乎

927 ~ 928

逻辑。而要超越这一点，就需要 1.4 节中提出的超出范围的概念。准确地说，下一个逻辑步骤是开发图像分析方法，这些方法的正确开发要求有来自于机器智能的概念。

如 1.1 节和 1.4 节中提到的那样，机器智能及其相关领域，譬如场景分析和计算机视觉，仍处于实际发展的早期阶段。今天，图像分析问题的求解仍然具有启发式描述的特征。尽管这些方法确实发生了变化，但它们的共同基础恰好是本书所覆盖的内容。

在学习了前 12 章的内容后，无论是从理论角度还是从实践角度，读者现在都应能理解数字图像处理所跨越的主要领域。请读者对所有的讨论细加琢磨，以便为以后深入研究该领域或以此为基础的相关领域打下坚实的基础。给定许多特定成像问题的性质后，清楚地理解基本原理可大大增加成功解决这些问题的机会。

参考文献

12.1 节到 12.2.2 节的背景材料源自 Theodoridis and Koutroumbas[2006]、Duda, Hart and Stork[2001]和 Tou and Gonzalez[1974]。Jain et al.[2000]撰写的综述性文章也很重要。Principe et al.[1999]很好地概述了神经网络。IEEE Trans. Image Processing[1998]的一期专刊可以与 10 年前的类似专刊相媲美(IEEE Computer[1988])。12.2.3 节中给出的材料是介绍性的。实际上，该讨论中使用的神经网络模型是多年来提出的大量模型中的一个。但我们论述的模型是有代表性的，而且是在图像处理过程中使用得相当广泛的模型。处理失真形状识别的例子改编自 Gupta et al.[1990, 1994]。Gori and Scarselli[1998]讨论了多层神经网络的分类能力。Udea[2000]报告的一个基于线性连接的神经网络的方法可使分类错误降至最低。

12.3.1 节中内容的其他读物，见 Bribiesca and Guzman[1980]。关于串匹配的读物，见 Sze and Yang[1981]、Oommen and Loke[1997]和 Gdalyahu and Weinshall[1999]。结构模式识别的参考材料见 Gonzalez and Thomason[1978]、Fu[1982]、Bunke and Sanfeliu[1990]、Tanaka[1995]、Vailaya et al.[1998]、Aizaka and Nakamura [1999]和 Jonk et al.[1999]，也可参见 Huang[2002]。

习题

标有星号的习题的详细答案，可在本书的网站上找到。该网站还基于本章的内容给出了建议的项目。

12.1 (a) 对图 12.1 中所示的模式，计算最小距离分类器的决策函数。通过仔细观察，你可能会得到所需的均值向量。

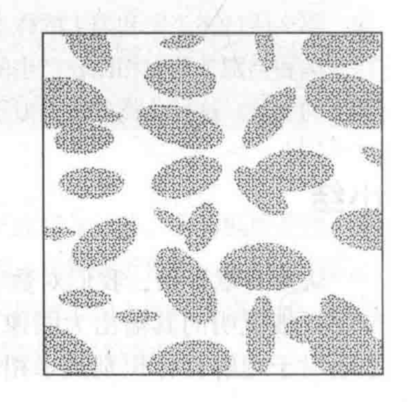
(b) 画出由(a)中决策函数实现的决策面。

★12.2 根据模式分类，说明式(12.2-4)和式(12.2-5)的功能相同。

12.3 证明由式(12.2-6)给出的面是连接 n 维点 $\mathbf{m}_i$ 和 $\mathbf{m}_j$ 的直线的垂直平分线。

★12.4 联系对图 12.7 的讨论，说明最小距离分类器是如何使用 W 组电阻(W 为类别的数量)，在每组处做求和连接(对电流求和)和一个可选取 W 个输入中最大输入的最大选择器来实现的，这里输入为电流。

12.5 证明式(12.2-8)的相关系数的值域为[−1, 1]。[提示：以向量形式表示 $\gamma(x, y)$。]



★12.6 某实验产生了斑点的二值图像，这些斑点的形态近似为椭圆(见右图)。斑点有三种尺寸，这些椭圆的主轴的平均值为(1.3, 0.7), (1.0, 0.5), (0.75, 0.25)。这些轴的尺寸相对于它们的平均值

928
~
929

有 ± 10%的变化。请开发一个图像处理系统，要求该系统能够拒绝不完整的或相互重叠的椭圆，然后将剩下的单个椭圆分类到给定的三种尺寸类之一中去。请以方框图的形式说明你的解决方案，对于每个框图的操作给出详尽的细节。使用最小距离分类器解决分类问题,，清楚地指出如何获得训练样本，以及如何使用这些样本来训练分类器。

12.7 下列模式类具有高斯概率密度函数：ω_1: $\{(0,0)^T,(4,0)^T,(4,4)^T,(0,4)^T\}$和$\omega_2$: $\{(5,5)^T,(7,5)^T,(7,7)^T,(5,7)^T\}$。

(a) 假设 $P(\omega_1)=P(\omega_2)=1/4$，请给出这两个类间的贝叶斯决策边界的方程。

(b) 画出该边界。

★12.8 重复习题 12.7，但使用下列模式类：ω_1: $\{(-1,0)^T,(0,-1)^T,(1,0)^T,(0,1)^T\}$和$\omega_2$:$\{(-4,0)^T,(0,-4)^T,(4,0)^T,(0,4)^T\}$。显然，这些类不是线性可分的。

12.9 重复习题 12.6，但使用一个贝叶斯分类器(假设其是高斯密度分布的)。请清楚地指出你是如何得到训练样本的，以及你是如何使用这些样本来训练分类器的。

★12.10 贝叶斯决策函数 $d_j(\boldsymbol{x})=p(\boldsymbol{x}/\omega_j)P(\omega_j), j=1,2,\cdots,W$ 是使用一个 0-1 损失函数推导的。证明这些决策函数可最小化错误分类的概率。[提示：错误分类概率 $p(e)$是 $1-p(c)$，其中 $p(c)$是正确分类的概率。对于一个属于类ω_i的模式向量 $\boldsymbol{x}$，有 $p(c/\boldsymbol{x})=p(\omega_i/\boldsymbol{x})$。求 $p(c)$，并证明当 $p(\boldsymbol{x}/\omega_i)P(\omega_i)$是最大值时 $p(c)$ 也是最大值(而 $p(e)$为最小值)]。

12.11 (a) 对下列模式类应用感知机算法：ω_1: $\{(0,0,0)^T,(1,0,0)^T,(1,0,1)^T,(1,1,0)^T\}$和$\omega_2$: $\{(0,0,1)^T,(0,1,1)^T,(0,1,0)^T,(1,1,1)^T\}$。令 $c=1$ 且 $\boldsymbol{w}(1)=(-1,-1,-2,0)^T$。

(b) 画出(a)中得到的决策面。给出这些模式类并指出该决策面的正侧。

★12.12 将类ω_2 的模式乘以 -1，可以按一种更为简洁的形式来表达式(12.2-34)到式(12.2-36)中给出的感知机算法；此时，算法中的修正步骤变为：若 $\boldsymbol{w}^T(k)\boldsymbol{y}(k)>0$，则 $\boldsymbol{w}(k+1)=\boldsymbol{w}(k)$，否则 $\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+c\boldsymbol{y}(k)$。这是从下面的通用梯度下降公式开始，推导出来的几个感知机算法公式之一：

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)-c\left[\frac{\partial J(\boldsymbol{w},\boldsymbol{y})}{\partial \boldsymbol{w}}\right]_{\boldsymbol{w}=\boldsymbol{w}(k)}$$

其中 $c>0$, $J(\boldsymbol{w},\boldsymbol{y})$是一个准则函数，偏导数在 $\boldsymbol{w}=\boldsymbol{w}(k)$处计算。证明使用准则函数 $J(\boldsymbol{w},\boldsymbol{y})=1/2(|\boldsymbol{w}^T\boldsymbol{y}|-\boldsymbol{w}^T\boldsymbol{y})$，可由通用梯度下降过程得到该感知机算法公式，其中|arg|是宗量的绝对值。(注意：$\boldsymbol{w}^T\boldsymbol{y}$ 关于 $\boldsymbol{w}$ 的偏导数等于 $\boldsymbol{y}$。)

12.13 证明：如果训练模式集合线性可分，那么式(12.2-34)到式(12.2-36)中给出的感知机训练算法会在有限步内收敛。[提示：将类ω_2的模式乘以 -1，并考虑一个非负阈值 T，以便感知机训练算法($c=1$)可表示为：若 $\boldsymbol{w}^T(k)\boldsymbol{y}(k)>T$，则 $\boldsymbol{w}(k+1)=\boldsymbol{w}(k)$；否则 $\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\boldsymbol{y}(k)$。你可能需要用到柯西-施瓦茨不等式$\|\boldsymbol{a}\|^2\|\boldsymbol{b}\|^2\geqslant(\boldsymbol{a}^T\boldsymbol{b})^2$。]

★12.14 指定一个神经网络的结构和权重，该神经网络的功能完全与 n 维空间中两个模式类的最小距离分类器的功能相同。

12.15 指定一个神经网络的结构和权重，该神经网络的功能完全与 n 维空间中两个模式类的贝叶斯分类器的功能相同。这些类是高斯型的，具有不同的均值，但协方差矩阵相等。

★12.16 (a) 在什么条件下，习题 12.14 和习题 12.15 中的神经网络相同?

(b) 如果使用足够数量的样本进行训练，那么用于12.2.3 节中讨论的多层前馈神经网络的通用Dleta准则会生成(a)中的特殊神经网络吗?

12.17 在两个维度上的两个模式类按如下方式分布：类ω_1的模式沿半径为 r_1 的一个圆随机分布。类似地，类ω_2的模式沿半径为 r_2 的一个圆随机地分布，其中 $r_2=4r_1$。为正确地分类这两类模式，请指定一个具有最少数量的层和节点的神经网络。

930 ~ 931

★12.18 重复习题 12.6，但使用一个神经网络。请清楚地指出如何得到训练样本，并说明如何使用这些样本来训练分类器。请按你的主张选择可解决该问题的最简神经网络。

12.19 证明式(12.2-71)中给出的表达式$h'_j(I_j)=O_j(1-O_j)$是由式(12.2-50)在$\theta_o=1$时推导出来的，其中$h'_j(I_j)=\partial h_j(I_j)/\partial I_j$。

★12.20 证明式(12.3-2)中的距离测度$D(A,B)$满足式(12.3-3)给出的性质。

12.21 说明当且仅当a和b是相同的串时，式(12.3-4)中的$\beta=\max(|a|,|b|)-\alpha$为 0。

12.22 某工厂为运动会生产美国国旗。质量保证小组发现，在生产高峰期间，某些印刷机器有(随机地)少印 1～3 颗星或 1～2 条条纹的倾向。除了这些错误之外，在其他方面国旗都很完美。尽管对于全部产品来说，含有错误的国旗的数量很少，但工厂经理仍决定解决该问题。仔细研究后，他认为使用图像处理技术来进行自动检测是解决该问题的最经济方法。基本规格如下：国旗大小约为 7.5 cm × 12.5 cm。他们将生产线的纵向移动(独立地，但方向上有 ±10° 变化)速度减小到约 40 cm/s，国旗之间的间隔约为 5 cm。所有情形下，“约”都意味着 ±5%。假设工厂经理雇用你为生产线设计一个图像处理系统。在论证方法的可行性时，你被告知成本和简单性是很重要的考虑因素。请基于图 1.23 设计一个完整的系统。将你的解决方案(包括各种假设和规格)简单(但清楚)地写成文档并上报给工厂经理。



附录 A　图像压缩编码表

引言

本附录包含了 CCITT 和 JPEG 压缩使用的码表。表 A.1 和表 A.2 是 CCITT Group 3 和 CCITT Group 4 压缩的改进霍夫曼编码。表 A.3 至表 A.5 是 JPEG DCT 系数的编码。使用这些表的详细信息，请参阅第 8 章中的 8.2.5 节和 8.2.8 节。

表 A.1　CCITT 终结码

行程	白色码字	黑色码字	行程	白色码字	黑色码字
0	00110101	0000110111	32	00011011	000001101010
1	000111	010	33	00010010	000001101011
2	0111	11	34	00010011	000011010010
3	1000	10	35	00010100	000011010011
4	1011	011	36	00010101	000011010100
5	1100	0011	37	00010110	000011010101
6	1110	0010	38	00010111	000011010110
7	1111	00011	39	00101000	000011010111
8	10011	000101	40	00101001	000001101100
9	10100	000100	41	00101010	000001101101
10	00111	0000100	42	00101011	000011011010
11	01000	0000101	43	00101100	000011011011
12	001000	0000111	44	00101101	000001010100
13	000011	00000100	45	00000100	000001010101
14	110100	00000111	46	00000101	000001010110
15	110101	000011000	47	00001010	000001010111
16	101010	0000010111	48	00001011	000001100100
17	101011	0000011000	49	01010010	000001100101
18	0100111	0000001000	50	01010011	000001010010
19	0001100	00001100111	51	01010100	000001010011
20	0001000	00001101000	52	01010101	000000100100
21	0010111	00001101100	53	00100100	000000110111
22	0000011	00000110111	54	00100101	000000111000
23	0000100	00000101000	55	01011000	000000100111
24	0101000	00000010111	56	01011001	000000101000
25	0101011	00000011000	57	01011010	000001011000
26	0010011	000011001010	58	01011011	000001011001
27	0100100	000011001011	59	01001010	000000101011

(续表)

行程	白色码字	黑色码字	行程	白色码字	黑色码字
28	0011000	000011001100	60	01001011	000000101100
29	00000010	000011001101	61	00110010	000001011010
30	00000011	000001101000	62	00110011	000001100110
31	00011010	000001101001	63	00110100	000001100111

表 A.2 CCITT 补充码

行程	白色码字	黑色码字	行程	白色码字	黑色码字
64	11011	0000001111	960	011010100	0000001110011
128	10010	000011001000	1024	011010101	0000001110100
192	010111	000011001001	1088	011010110	0000001110101
256	0110111	000001011011	1152	011010111	0000001110110
320	00110110	000000110011	1216	011011000	0000001110111
384	00110111	000000110100	1280	011011001	0000001010010
448	01100100	000000110101	1344	011011010	0000001010011
512	01100101	0000001101100	1408	011011011	0000001010100
576	01101000	0000001101101	1472	010011000	0000001010101
640	01100111	0000001001010	1536	010011001	0000001011010
704	011001100	0000001001011	1600	010011010	0000001011011
768	011001101	0000001001100	1664	011000	0000001100100
832	011010010	0000001001101	1728	010011011	0000001100101
896	011010011	0000001110010			

	码字		码字
1792	00000001000	2240	000000010110
1856	00000001100	2304	000000010111
1920	00000001101	2368	000000011100
1984	000000010010	2432	000000011101
2048	000000010011	2496	000000011110
2112	000000010100	2560	000000011111
2176	000000010101		

表 A.3 JPEG 系数编码分类

值域	DC差值分类	AC分类
0	0	N/A
−1, 1	1	1
−3, −2, 2, 3	2	2
−7, ⋯, −4, 4, ⋯, 7	3	3
−15, ⋯, −8, 8, ⋯, 15	4	4
−31, ⋯, −16, 16, ⋯, 31	5	5
−63, ⋯, −32, 32, ⋯, 63	6	6
−127, ⋯, −64, 64, ⋯, 127	7	7
−255, ⋯, −128, 128, ⋯, 255	8	8
−511, ⋯, −256, 256, ⋯, 511	9	9
−1023, ⋯, −512, 512, ⋯, 1023	A	A
−2047, ⋯, −1024, 1024, ⋯, 2047	B	B
−4095, ⋯, −2048, 2048, ⋯, 4095	C	C
−8191, ⋯, −4096, 4096, ⋯, 8191	D	D
−16383, ⋯, −8192, 8192, ⋯, 16383	E	E
−32767, ⋯, −16384, 16384, ⋯, 32767	F	N/A

表 A.4 JPEG 默认的 DC 编码(照度)

类别	基码	长度	类别	基码	长度
0	010	3	6	1110	10
1	011	4	7	11110	12
2	100	5	8	111110	14
3	00	5	9	1111110	16
4	101	7	A	11111110	18
5	110	8	B	111111110	20

933 ~ 935

表 A.5　JPEG 默认的 AC 编码(照度)

行程/类别	基码	长度	行程/类别	基码	长度
0/0	**1010 (= EOB)**	**4**			
0/1	00	3	8/1	11111010	9
0/2	01	4	8/2	111111111000000	17
0/3	100	6	8/3	1111111110110111	19
0/4	1011	8	8/4	1111111110111000	20
0/5	11010	10	8/5	1111111110111001	21
0/6	111000	12	8/6	1111111110111010	22
0/7	1111000	14	8/7	1111111110111011	23
0/8	1111110110	18	8/8	1111111110111100	24
0/9	1111111110000010	25	8/9	1111111110111101	25
0/A	1111111110000011	26	8/A	1111111110111110	26
1/1	1100	5	9/1	111111000	10
1/2	111001	8	9/2	1111111110111111	18
1/3	1111001	10	9/3	1111111111000000	19
1/4	111110110	13	9/4	1111111111000001	20
1/5	11111110110	16	9/5	1111111111000010	21
1/6	1111111110000100	22	9/6	1111111111000011	22
1/7	1111111110000101	23	9/7	1111111111000100	23
1/8	1111111110000110	24	9/8	1111111111000101	24
1/9	1111111110000111	25	9/9	1111111111000110	25
1/A	1111111110001000	26	9/A	1111111111000111	26
2/1	11011	6	A/1	111111001	10
2/2	11111000	10	A/2	1111111111001000	18
2/3	1111110111	13	A/3	1111111111001001	19
2/4	1111111110001001	20	A/4	1111111111001010	20
2/5	1111111110001010	21	A/5	1111111111001011	21
2/6	1111111110001011	22	A/6	1111111111001100	22
2/7	1111111110001100	23	A/7	1111111111001101	23
2/8	1111111110001101	24	A/8	1111111111001110	24
2/9	1111111110001110	25	A/9	1111111111001111	25
2/A	1111111110001111	26	A/A	1111111111010000	26
3/1	111010	7	B/1	111111010	10
3/2	111110111	11	B/2	1111111111010001	18
3/3	11111110111	14	B/3	1111111111010010	19
3/4	1111111110010000	20	B/4	1111111111010011	20
3/5	1111111110010001	21	B/5	1111111111010100	21
3/6	1111111110010010	22	B/6	1111111111010101	22
3/7	1111111110010011	23	B/7	1111111111010110	23
3/8	1111111110010100	24	B/8	1111111111010111	24
3/9	1111111110010101	25	B/9	1111111111011000	25
3/A	1111111110010110	26	B/A	1111111111011001	26
4/1	111011	7	C/1	1111111010	11
4/2	1111111000	12	C/2	1111111111011010	18
4/3	1111111110010111	19	C/3	1111111111011011	19
4/4	1111111110011000	20	C/4	1111111111011100	20
4/5	1111111110011001	21	C/5	1111111111011101	21
4/6	1111111110011010	22	C/6	1111111111011110	22
4/7	1111111110011011	23	C/7	1111111111011111	23
4/8	1111111110011100	24	C/8	1111111111100000	24
4/9	1111111110011101	25	C/9	1111111111100001	25
4/A	1111111110011110	26	C/A	1111111111100010	26
5/1	1111010	8	D/1	11111111010	12
5/2	1111111001	12	D/2	1111111111100011	18
5/3	1111111110011111	19	D/3	1111111111100100	19
5/4	1111111110100000	20	D/4	1111111111100101	20
5/5	1111111110100001	21	D/5	1111111111100110	21
5/6	1111111110100010	22	D/6	1111111111100111	22
5/7	1111111110100011	23	D/7	1111111111101000	23
5/8	1111111110100100	24	D/8	1111111111101001	24
5/9	1111111110100101	25	D/9	1111111111101010	25
5/A	1111111110100110	26	D/A	1111111111101011	26
6/1	1111011	8	E/1	111111110110	13

935 ~ 936

(续表)

行程/类别	基码	长度	行程/类别	基码	长度
6/2	11111111000	13	E/2	1111111111101100	18
6/3	1111111110100111	19	E/3	1111111111101101	19
6/4	1111111110101000	20	E/4	1111111111101110	20
6/5	1111111110101001	21	E/5	1111111111101111	21
6/6	1111111110101010	22	E/6	1111111111110000	22
6/7	1111111110101011	23	E/7	1111111111110001	23
6/8	1111111110101100	24	E/8	1111111111110010	24
6/9	1111111110101101	25	E/9	1111111111110011	25
6/A	1111111110101110	26	E/A	1111111111110100	26
7/1	11111001	9	**F/0**	**111111110111**	**12**
7/2	11111111001	13	F/1	1111111111110101	17
7/3	1111111110101111	19	F/2	1111111111110110	18
7/4	1111111110110000	20	F/3	1111111111110111	19
7/5	1111111110110001	21	F/4	1111111111111000	20
7/6	1111111110110010	22	F/5	1111111111111001	21
7/7	1111111110110011	23	F/6	1111111111111010	22
7/8	1111111110110100	24	F/7	1111111111111011	23
7/9	1111111110110101	25	F/8	1111111111111100	24
7/A	1111111110110110	26	F/9	1111111111111101	25
			F/A	1111111111111110	26

附录B 参 考 书 目

Abidi, M. A. and Gonzalez, R. C. (eds.) [1992]. *Data Fusion in Robotics and Machine Intelligence*, Academic Press, New York.

Abidi, M. A., Eason, R. O., and Gonzalez, R. C. [1991]. "Autonomous Robotics Inspection and Manipulation Using Multisensor Feedback," *IEEE Computer*, vol. 24, no. 4, pp. 17–31.

Abramson, N. [1963]. *Information Theory and Coding*, McGraw-Hill, New York.

Adiv, G. [1985]. "Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 4, pp. 384–401.

Aggarwal, J. K. and Badler, N. I. (eds.) [1980]. "Motion and Time-Varying Imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, Special Issue, vol. PAMI-2, no. 6, pp. 493–588.

Aguado, A. S., Nixon, M. S., and Montiel, M. M. [1998], "Parameterizing Arbitrary Shapes via Fourier Descriptors for Evidence-Gathering Extraction," *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 202–221.

Ahmed, N., Natarajan, T., and Rao, K. R. [1974]. "Discrete Cosine Transforms," *IEEE Trans. Comp.*, vol. C-23, pp. 90–93.

Ahmed, N. and Rao, K. R. [1975]. *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag, New York.

Aizaka, K. and Nakamura, A. [1999]. "Parsing of Two-Dimensional Images Represented by Quadtree Adjoining Grammars," *Pattern Recog.*, vol. 32, no. 2, pp. 277–294.

Alexiadis, D. S. and Sergiadis, G. D. [2007]. "Estimation of Multiple Accelerated Motions Using Chirp-Fourier Transforms and Clustering," *IEEE Trans. Image Proc.*, vol. 16, no. 1, pp. 142–152.

Alliney, S. [1993]. "Digital Analysis of Rotated Images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 5, pp. 499–504.

Ando, S. [2000]. "Consistent Gradient Operators," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 3, pp. 252–265.

Andrews, H. C. [1970]. *Computer Techniques in Image Processing*, Academic Press, New York.

Andrews, H. C. and Hunt, B. R. [1977]. *Digital Image Restoration*, Prentice Hall, Englewood Cliffs, N.J.

Anelli, G., Broggi, A., and Destri, G. [1998]. "Decomposition of Arbitrarily-Shaped Morphological Structuring Elements Using Genetic Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 2, pp. 217–224.

Ang, P. H., Ruetz, P. A., and Auld, D. [1991]. "Video Compression Makes Big Gains," *IEEE Spectrum*, vol. 28, no. 10, pp. 16–19.

Antonini, M., Barlaud, M., Mathieu, P., and Daubechies, I. [1992]. "Image Coding Using Wavelet Transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220.

Ascher, R. N. and Nagy, G. [1974]. "A Means for Achieving a High Degree of Compaction on Scan-Digitized Printed Text," *IEEE Transactions on Comp.*, C-23:1174–1179.

Atchison, D. A. and Smith, G. [2000]. *Optics of the Human Eye*, Butterworth-Heinemann, Boston, Mass.

Baccar, M., Gee, L. A., Abidi, M. A., and Gonzalez, R. C. [1996]. "Segmentation of Range Images Via Data Fusion and Morphological Watersheds," *Pattern Recog.*, vol. 29, no. 10, pp. 1671–1685.

Bajcsy, R. and Lieberman, L. [1976]. "Texture Gradient as a Depth Cue," *Comput. Graph. Image Proc.*, vol. 5, no. 1, pp. 52–67.

Bakir, T. and Reeves, J. S. [2000]. "A Filter Design Method for Minimizing Ringing in a

Region of Interest in MR Spectroscopic Images," *IEEE Trans. Medical Imaging*, vol. 19, no. 6, pp. 585–600.

Ballard, D. H. [1981]. "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122.

Ballard, D. H. and Brown, C. M. [1982]. *Computer Vision*, Prentice Hall, Englewood Cliffs, N.J.

Banham, M. R., Galatsanos, H. L., Gonzalez, H. L., and Katsaggelos, A. K. [1994]. "Multichannel Restoration of Single Channel Images Using a Wavelet-Based Subband Decomposition," *IEEE Trans. Image Processing*, vol. 3, no. 6, pp. 821–833.

Banham, M. R. and Katsaggelos, A. K. [1996]. "Spatially Adaptive Wavelet-Based Multiscale Image Restoration," *IEEE Trans. Image Processing*, vol. 5, no. 5, pp. 619–634.

Basart, J. P. and Gonzalez, R. C. [1992]. "Binary Morphology," in *Advances in Image Analysis*, Y. Mahdavieh and R. C. Gonzalez (eds.), SPIE Press, Bellingham, Wash., pp. 277–305.

Basart, J. P., Chacklackal, M. S., and Gonzalez, R. C. [1992]. "Introduction to Gray-Scale Morphology," in *Advances in Image Analysis*, Y. Mahdavieh and R. C. Gonzalez (eds.), SPIE Press, Bellingham, Wash., pp. 306–354.

Bates, R. H. T. and McDonnell, M. J. [1986]. *Image Restoration and Reconstruction*, Oxford University Press, New York.

Battle, G. [1987]. "A Block Spin Construction of Ondelettes. Part I: Lemarié Functions," *Commun. Math. Phys.*, vol. 110, pp. 601–615.

Battle, G. [1988]. "A Block Spin Construction of Ondelettes. Part II: the QFT Connection," *Commun. Math. Phys.*, vol. 114, pp. 93–102.

Baumert, L. D., Golomb, S. W., and Hall, M., Jr. [1962]. "Discovery of a Hadamard Matrix of Order 92," *Bull. Am. Math. Soc.*, vol. 68, pp. 237–238.

Baxes, G. A. [1994]. *Digital Image Processing: Principles and Applications*, John Wiley & Sons, New York.

Baylon, D. M. and Lim, J. S. [1990]. "Transform/Subband Analysis and Synthesis of Signals," *Tech. Report*, MIT Research Laboratory of Electronics, Cambridge, Mass.

Bell, E. T. [1965]. *Men of Mathematics*, Simon & Schuster, New York.

Bengtsson, A. and Eklundh, J. O. [1991]. "Shape Representation by Multiscale Contour Approximation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 1, pp. 85–93.

Benson, K. B. [1985]. *Television Engineering Handbook*, McGraw-Hill, New York.

Berger, T. [1971]. *Rate Distortion Theory*, Prentice Hall, Englewood Cliffs, N.J.

Beucher, S. [1990]. Doctoral Thesis, Centre de Morphologie Mathématique, École des Mines de Paris, France. (The core of this material is contained in the following paper.)

Beucher, S. and Meyer, F. [1992]. "The Morphological Approach of Segmentation: The Watershed Transformation," in *Mathematical Morphology in Image Processing*, E. Dougherty (ed.), Marcel Dekker, New York.

Bezdek, J. C., et al. [2005]. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer, New York.

Bhaskaran, V. and Konstantinos, K. [1997]. *Image and Video Compression Standards: Algorithms and Architectures*, Kluwer, Boston, Mass.

Bhatt, B., Birks, D., Hermreck, D. [1997]. "Digital Television: Making It Work," *IEEE Spectrum*, vol. 34, no. 10, pp. 19–28.

Biberman, L. M. [1973]. "Image Quality," In *Perception of Displayed Information*, Biberman, L. M. (ed.), Plenum Press, New York.

Bichsel, M. [1998]. "Analyzing a Scene's Picture Set under Varying Lighting," *Computer Vision and Image Understanding*, vol. 71, no. 3, pp. 271–280.

Bieniek, A. and Moga, A. [2000]. "An Efficient Watershed Algorithm Based on Connected Components," *Pattern Recog.*, vol. 33, no. 6, pp. 907–916.

Bisignani, W. T., Richards, G. P., and Whelan, J. W. [1966]. "The Improved Grey Scale and Coarse-Fine PCM Systems: Two New Digital TV Bandwidth Reduction Techniques," *Proc. IEEE*, vol. 54, no. 3, pp. 376–390.

Blahut, R. E. [1987]. *Principles and Practice of Information Theory*, Addison-Wesley, Reading, Mass.

Bleau, A. and Leon, L. J. [2000]. "Watershed-Based Segmentation and Region Merging," *Computer Vision and Image Understanding*, vol. 77, no. 3, pp. 317–370.

Blouke, M. M., Sampat, N., and Canosa, J. [2001]. *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications-II*, SPIE Press, Bellingham, Wash.

Blum, H. [1967]. "A Transformation for Extracting New Descriptors of Shape," In *Models for the Perception of Speech and Visual Form*, Wathen-Dunn, W. (ed.), MIT Press, Cambridge, Mass.

Blume, H. and Fand, A. [1989]. "Reversible and Irreversible Image Data Compression Using the S-Transform and Lempel-Ziv Coding," *Proc. SPIE Medical Imaging III: Image Capture and Display*, vol. 1091, pp. 2–18.

Boie, R. A. and Cox, I. J. [1992]. "An Analysis of Camera Noise," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 6, pp. 671–674.

Born, M. and Wolf, E. [1999]. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed., Cambridge University Press, Cambridge, UK.

Boulgouris, N. V., Tzovaras, D., and Strintzis, M. G. [2001]. "Lossless Image Compression Based on Optimal Prediction, Adaptive Lifting, and Conditional Arithmetic Coding," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 1–14.

Bouman, C. and Liu, B. [1991]. "Multiple Resolution Segmentation of Textured Images," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. 13, no. 2, pp. 99–113.

Boyd, J. E. and Meloche, J. [1998]. "Binary Restoration of Thin Objects in Multidimensional Imagery," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 6, pp. 647–651.

Bracewell, R. N. [1995]. *Two-Dimensional Imaging*, Prentice Hall, Upper Saddle River, N.J.

Bracewell, R. N. [2000]. *The Fourier Transform and its Applications*, 3rd ed. McGraw-Hill, New York.

Brechet, L., Lucas, M., Doncarli, C., and Farnia, D. [2007]. "Compression of Biomedical Signals with Mother Wavelet Optimization and Best-Basis Wavelet Packet Selection," *IEEE Trans. on Biomedical Engineering*, in press.

Bribiesca, E. [1981]. "Arithmetic Operations Among Shapes Using Shape Numbers," *Pattern Recog.*, vol. 13, no. 2, pp. 123–138.

Bribiesca, E. [1999]. "A New Chain Code," *Pattern Recog.*, vol. 32, no. 2, pp. 235–251.

Bribiesca, E. [2000]. "A Chain Code for Representing 3–D Curves," *Pattern Recog.*, vol. 33, no. 5, pp. 755–765.

Bribiesca, E. and Guzman, A. [1980]. "How to Describe Pure Form and How to Measure Differences in Shape Using Shape Numbers," *Pattern Recog.*, vol. 12, no. 2, pp. 101–112.

Brigham, E. O. [1988]. *The Fast Fourier Transform and its Applications*, Prentice Hall, Upper Saddle River, N.J.

Brinkman, B. H., Manduca, A., and Robb, R. A. [1998]. "Optimized Homomorphic Unsharp Masking for MR Grayscale Inhomogeneity Correction," *IEEE Trans. Medical Imaging*, vol. 17, no. 2, pp. 161–171.

Brummer, M. E. [1991]. "Hough Transform Detection of the Longitudinal Fissure in Tomographic Head Images," *IEEE Trans. Biomed. Images*, vol. 10, no. 1, pp. 74–83.

Brzakovic, D., Patton, R., and Wang, R. [1991]. "Rule-Based Multi-Template Edge Detection," *Comput. Vision, Graphics, Image Proc: Graphical Models and Image Proc.*, vol. 53, no. 3, pp. 258–268.

Bunke, H. and Sanfeliu, A. (eds.) [1990]. *Syntactic and Structural Pattern Recognition: Theory and Applications*, World Scientific, Teaneck, N.J.

Burrus, C. S., Gopinath, R. A., and Guo, H. [1998]. *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, Upper Saddle River, N.J., pp. 250–251.

Burt, P. J. and Adelson, E. H. [1983]. "The Laplacian Pyramid as a Compact Image Code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540.

Cameron, J. P. [2005]. *Sets, Logic, and Categories*, Springer, New York.

Campbell, J. D. [1969]. "Edge Structure and the Representation of Pictures," Ph.D. dissertation, Dept. of Elec. Eng., University of Missouri, Columbia.

Candy, J. C., Franke, M. A., Haskell, B. G., and Mounts, F. W. [1971]. "Transmitting

Television as Clusters of Frame-to-Frame Differences," *Bell Sys. Tech. J.*, vol. 50, pp. 1889–1919.
Cannon, T. M. [1974]. "Digital Image Deblurring by Non-Linear Homomorphic Filtering," Ph.D. thesis, University of Utah.
Canny, J. [1986]. "A Computational Approach for Edge Detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679–698.
Carey, W. K., Chuang, D. B., and Hamami, S. S. [1999]. "Regularity-Preserving Image Interpolation," *IEEE Trans. Image Processing*, vol. 8, no. 9, pp. 1293–1299.
Caselles, V., Lisani, J.-L., Morel, J.-M., and Sapiro, G. [1999]. "Shape Preserving Local Histogram Modification,"*IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 220–230.
Castleman, K. R. [1996]. *Digital Image Processing*, 2nd ed. Prentice Hall, Upper Saddle River, N.J.
Centeno, J. A. S. and Haertel, V. [1997]. "An Adaptive Image Enhancement Algorithm," *Pattern Recog.*, vol. 30, no. 7 pp. 1183–1189.
Chan, R. C., Karl, W. C., and Lees, R. S. [2000]. "A New Model-Based Technique for Enhanced Small-Vessell Measurements in X-Ray Cine-Angiograms," *IEEE Trans. Medical Imaging*, vol. 19, no. 3, pp. 243–255.
Chandler, D. and Hemami, S. [2005]. "Dynamic Contrast-Based Quantization for Lossy Wavelet Image Compression," *IEEE Trans. Image Proc.*, vol. 14, no. 4, pp. 397–410.
Chang, S. G., Yu, B., and Vetterli, M. [2000]. "Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1522–1531.
Chang, S. K. [1989]. *Principles of Pictorial Information Systems Design*, Prentice Hall, Englewood Cliffs, N.J.
Chang, T. and Kuo, C.-C. J. [1993]. "Texture Analysis and Classification with Tree-Structures Wavelet Transforms," *IEEE Trans. Image Processing*, vol. 2, no. 4, pp. 429–441.
Champeney, D. C. [1987]. *A Handbook of Fourier Theorems*, Cambridge University Press, New York.
Chaudhuri, B. B. [1983]. "A Note on Fast Algorithms for Spatial Domain Techniques in Image Processing," *IEEE Trans. Syst. Man Cyb.*, vol. SMC-13, no. 6, pp. 1166–1169.
Chen, M. C. and Wilson, A. N. [2000]. "Motion-Vector Optimization of Control Grid Interpolation and Overlapped Block Motion Compensation Using Iterated Dynamic Programming," *IEEE Trans. Image Processing*., vol. 9, no. 7, pp. 1145–1157.
Chen, Y.-S. and Yu, Y.-T. [1996]. "Thinning Approach for Noisy Digital Patterns," *Pattern Recog.*, vol. 29, no. 11, pp. 1847–1862.
Cheng, H. D. and Huijuan Xu, H. [2000]. "A Novel Fuzzy Logic Approach to Contrast Enhancement," *Pattern Recog.*, vol. 33, no. 5, pp. 809–819.
Cheriet, M., Said, J. N., and Suen, C. Y. [1998]. "A Recursive Thresholding Technique for Image Segmentation," *IEEE Trans. Image Processing*, vol. 7, no. 6, pp. 918–921.
Cheung, J., Ferris, D., and Kurz, L. [1997]. "On Classification of Multispectral Infrared Image Data," *IEEE Trans. Image Processing*, vol. 6, no. 10, pp. 1456–1464.
Cheung, K. K. T. and Teoh, E. K. [1999]. "Symmetry Detection by Generalized Complex (GC) Moments: A Closed-Form Solution," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 5, pp. 466–476.
Chow, C. K. and Kaneko, T. [1972]. "Automatic Boundary Detection of the Left Ventricle from Cineangiograms," *Comp., and Biomed. Res.*, vol. 5, pp. 388–410.
Chu, C.-C. and Aggarwal, J. K. [1993]. "The Integration of Image Segmentation Maps Using Regions and Edge Information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 12, pp. 1241–1252.
CIE [1978]. *Uniform Color Spaces—Color Difference Equations—Psychometric Color Terms*, Commission Internationale de L'Eclairage, Publication No. 15, Supplement No. 2, Paris.
Clark, J. J. [1989]. "Authenticating Edges Produced by Zero-Crossing Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 8, pp. 830–831.
Clarke, R. J. [1985]. *Transform Coding of Images*, Academic Press, New York.

Cochran, W. T., Cooley, J. W., et al. [1967]. "What Is the Fast Fourier Transform?" *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 45–55.
Coeurjolly, D. and Klette, R. [2004]. "A Comparative Evaluation of Length Estimators of Digital Curves," *IEEE Trans. Pattern. Analysis Machine Int.*, vol. 26, no. XX, pp. 252–258.
Cohen, A. and Daubechies, I. [1992]. *A Stability Criterion for Biorthogonal Wavelet Bases and Their Related Subband Coding Schemes*, Technical Report, AT&T Bell Laboratories.
Cohen, A., Daubechies, I., and Feauveau, J.-C. [1992]. "Biorthogonal Bases of Compactly Supported Wavelets," *Commun. Pure and Appl. Math.*, vol. 45, pp. 485–560.
Coifman, R. R. and Wickerhauser, M. V. [1992]. "Entropy-Based Algorithms for Best Basis Selection," *IEEE Tran. Information Theory*, vol. 38, no. 2, pp. 713–718.
Coltuc, D., Bolon, P., and Chassery, J. M. [2006]. "Exact Histogram Specification," *IEEE Trans. Image Processing*, vol. 15, no. 5, pp. 1143–1152.
Cooley, J. W., Lewis, P. A. W., and Welch, P. D. [1967a]. "Historical Notes on the Fast Fourier Transform," *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 76–79.
Cooley, J. W., Lewis, P. A. W., and Welch, P. D. [1967b]. "Application of the Fast Fourier Transform to Computation of Fourier Integrals," *IEEE Trans. Audio and Electroacoustics*, vol. AU-15, no. 2, pp. 79–84.
Cooley, J. W., Lewis, P. A. W., and Welch, P. D. [1969]. "The Fast Fourier Transform and Its Applications," *IEEE Trans. Educ.*, vol. E-12, no. 1. pp. 27–34.
Cooley, J. W. and Tukey, J. W. [1965]. "An Algorithm for the Machine Calculation of Complex Fourier Series," *Math. of Comput.*, vol. 19, pp. 297–301.
Cornsweet, T. N. [1970]. *Visual Perception*, Academic Press, New York.
Cortelazzo, G. M., Lucchese, L., and Monti, C. M. [1999]. "Frequency Domain Analysis of General Planar Rigid Motion with Finite Duration," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 16, no. 6, pp. 1238–1253.
Cowart, A. E., Snyder, W. E., and Ruedger, W. H. [1983]. "The Detection of Unresolved Targets Using the Hough Transform," *Comput. Vision Graph Image Proc.*, vol. 21, pp. 222–238.
Cox, I., Kilian, J., Leighton, F., and Shamoon, T. [1997]. "Secure Spread Spectrum Watermarking for Multimedia," *IEEE Trans. Image Proc.*, vol. 6, no. 12, pp. 1673–1687.
Cox, I., Miller, M., and Bloom, J. [2001]. *Digital Watermarking*, Morgan Kaufmann (Elsevier), New York.
Creath, K. and Wyant, J. C. [1992]. "Moire and Fringe Patterns," in *Optical Shop Testing*, 2nd ed., (D. Malacara, ed.), John Wiley & Sons, New York, pp. 653–685.
Croisier, A., Esteban, D., and Galand, C. [1976]. "Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques," *Int. Conf. On Inform. Sciences and Systems*, Patras, Greece, pp. 443–446.
Cumani, A., Guiducci, A., and Grattoni, P. [1991]. "Image Description of Dynamic Scenes," *Pattern Recog.*, vol. 24, no. 7, pp. 661–674.
Cutrona, L. J. and Hall, W. D. [1968]. "Some Considerations in Post-Facto Blur Removal," In *Evaluation of Motion-Degraded Images*, NASA Publ. SP-193, pp. 139–148.
Danielson, G. C. and Lanczos, C. [1942]. "Some Improvements in Practical Fourier Analysis and Their Application to X-Ray Scattering from Liquids," *J. Franklin Institute*, vol. 233, pp. 365–380, 435–452.
Daubechies, I. [1988]. "Orthonormal Bases of Compactly Supported Wavelets," *Commun. On Pure and Appl. Math.*, vol. 41, pp. 909–996.
Daubechies, I. [1990]. "The Wavelet Transform, Time-Frequency Localization and Signal Analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005.
Daubechies, I. [1992]. *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pa.
Daubechies, I. [1993]. "Orthonormal Bases of Compactly Supported Wavelets II, Variations on a Theme," *SIAM J. Mathematical Analysis*, vol. 24, no. 2, pp. 499–519.
Daubechies, I. [1996]. "Where Do We Go from Here?—A Personal Point of View,"

Proc. IEEE, vol. 84, no. 4, pp. 510–513.

Daul, C., Graebling, P., and Hirsch, E. [1998]. "From the Hough Transform to a New Approach for the Detection of and Approximation of Elliptical Arcs," *Computer Vision and Image Understanding*, vol. 72, no. 3, pp. 215–236.

Davies, E. R. [2005]. *Machine Vision: Theory, Algorithms, Practicalities*, Morgan Kaufmann, San Francisco.

Davis, L. S. [1982]. "Hierarchical Generalized Hough Transforms and Line-Segment Based Generalized Hough Transforms," *Pattern Recog.*, vol. 15, no. 4, pp. 277–285.

Davis, T. J. [1999]. "Fast Decomposition of Digital Curves into Polygons Using the Haar Transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 8, pp. 786–790.

Davisson, L. D. [1972]. "Rate-Distortion Theory and Application," *Proc. IEEE*, vol. 60, pp. 800–808.

Delaney, A. H. and Bresler, Y. [1995]. "Multiresolution Tomographic Reconstruction Using Wavelets," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 799–813.

Delon, J., Desolneux, A., Lisani, J. L., and Petro, A. B. [2007]. "A Nonparametric Approach for Histogram Segmentation," *IEEE Trans. Image Proc.*, vol. 16, no. 1, pp. 253–261.

Delp, E. J. and Mitchell, O. R. [1979]. "Image Truncation using Block Truncation Coding," *IEEE Trans. Comm.*, vol. COM-27, pp. 1335–1342.

Di Zenzo, S. [1986]. "A Note on the Gradient of a Multi-Image," *Computer Vision, Graphics, and Image Processing*, vol. 33, pp. 116–125.

Dijkstra, E. [1959]. "Note on Two Problems in Connection with Graphs," *Numerische Mathematik*, vol. 1, pp. 269–271.

Djeziri, S., Nouboud, F., and Plamondon, R. [1998]. "Extraction of Signatures from Check Background Based on a Filiformity Criterion," *IEEE Trans. Image Processing*, vol. 7, no. 102, pp. 1425–1438.

Dougherty, E. R. [1992]. *An Introduction to Morphological Image Processing*, SPIE Press, Bellingham, Wash.

Dougherty, E. R. (ed.) [2000]. *Random Processes for Image and Signal Processing*, IEEE Press, New York.

Dougherty. E. R. and Lotufo, R. A. [2003]. *Hands-on Morphological Image Processing*, SPIE Press, Bellingham, WA.

Drew, M. S., Wei, J., and Li, Z.-N. [1999]. "Illumination Invariant Image Retrieval and Video Segmentation," *Pattern Recog.*, vol. 32, no. 8, pp. 1369–1388.

Duda, R. O. and Hart, P. E. [1972]. "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11–15.

Duda, R. O, Hart, P. E., and Stork, D. G. [2001]. *Pattern Classification*, John Wiley & Sons, New York.

Dugelay, J., Roche, S., Rey, C., and Doerr, G. [2006]. "Still-Image Watermarking Robust to Local Geometric Distortions," *IEEE Trans. Image Proc.*, vol. 15, no. 9, pp. 2831–2842.

Edelman, S. [1999]. *Representation and Recognition in Vision*, The MIT Press, Cambridge, Mass.

Elias, P. [1952]. "Fourier Treatment of Optical Processes," *J. Opt. Soc. Am.*, vol. 42, no. 2, pp. 127–134.

Elliott, D. F. and Rao, K. R. [1983]. *Fast Transforms: Algorithms and Applications*, Academic Press, New York.

Eng, H.-L. and Ma, K.-K. [2001]. "Noise Adaptive Soft-Switching Median Filter," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 242–251.

Eng, H.-L. and Ma, K.-K. [2006]. "A Switching Median Filter With Boundary Discriminitative Noise Detection for Extremely Corrupted Images," *IEEE Trans. Image Proc.*, vol. 15, no. 6, pp. 1506–1516.

Equitz, W. H. [1989]. "A New Vector Quantization Clustering Algorithm," *IEEE Trans. Acous. Speech Signal Processing*, vol. ASSP-37, no. 10, pp. 1568–1575.

Etienne, E. K. and Nachtegael, M. (eds.) [2000]. *Fuzzy Techniques in Image Processing*, Springer-Verlag, New York.

Evans, A. N. and Liu, X. U. [2006]. "A Morphological Gradient Approach to Color Edge Detection," *IEEE Trans. Image Proc.*, vol. 15, no. 6, pp. 1454–1463.

Falconer, D. G. [1970]. "Image Enhancement and Film Grain Noise." *Opt. Acta*, vol. 17, pp. 693–705.
Fairchild, M. D. [1998]. *Color Appearance Models*, Addison-Wesley, Reading, Mass.
Federal Bureau of Investigation [1993]. *WSQ Gray-Scale Fingerprint Image Compression Specification*, IAFIS-IC-0110v2, Washington, D. C.
Felsen, L. B. and Marcuvitz, N. [1994]. *Radiation and Scattering of Waves*, IEEE Press, New York.
Ferreira, A. and Ubéda, S. [1999]. "Computing the Medial Axis Transform in Parallel with Eight Scan Operators," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 3, pp. 277–282.
Fischler, M. A. [1980]. "Fast Algorithms for Two Maximal Distance Problems with Applications to Image Analysis," *Pattern Recog.*, vol. 12, pp. 35–40.
Fisher, R. A. [1936]. "The Use of Multiple Measurements in Taxonomic Problems," *Ann. Eugenics*, vol. 7, Part 2, pp. 179–188. (Also in *Contributions to Mathematical Statistics*, John Wiley & Sons, New York, 1950.)
Flusser, J. [2000]. "On the Independence of Rotation Moment Invariants," *Pattern Recog.*, vol. 33, pp. 1405–1410.
Forsyth, D. F. and Ponce, J. [2002]. *Computer Vision—A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.
Fortner, B. and Meyer, T. E. [1997]. *Number by Colors*, Springer-Verlag, New York.
Fox, E. A. [1991]. "Advances in Interactive Digital Multimedia Systems," *Computer*, vol. 24, no. 10, pp. 9–21.
Fram, J. R. and Deutsch, E. S. [1975]. "On the Quantitative Evaluation of Edge Detection Schemes and Their Comparison with Human Performance," *IEEE Trans. Computers*, vol. C-24, no. 6, pp. 616–628.
Freeman, A. (translator) [1878]. J. Fourier, *The Analytical Theory of Heat*, Cambridge: University Press, London.
Freeman, C. [1987]. *Imaging Sensors and Displays*, ISBN 0–89252–800–1, SPIE Press, Bellingham, Wash.
Freeman, H. [1961]. "On the Encoding of Arbitrary Geometric Configurations," *IEEE Trans. Elec. Computers*, vol. EC-10, pp. 260–268.
Freeman, H. [1974]. "Computer Processing of Line Drawings," *Comput. Surveys*, vol. 6, pp. 57–97.
Freeman, H. and Shapira, R. [1975]. "Determining the Minimum-Area Encasing Rectangle for an Arbitrary Closed Curve," *Comm. ACM*, vol. 18, no. 7, pp. 409–413.
Freeman, J. A. and Skapura, D. M. [1991]. *Neural Networks: Algorithms, Applications, and Programming Techniques*, Addison-Wesley, Reading, Mass.
Frei, W. and Chen, C. C. [1977]. "Fast Boundary Detection: A Generalization and a New Algorithm," *IEEE Trans. Computers*, vol. C-26, no. 10, pp. 988–998.
Frendendall, G. L. and Behrend, W. L. [1960]. "Picture Quality—Procedures for Evaluating Subjective Effects of Interference," *Proc. IRE*, vol. 48, pp. 1030–1034.
Fu, K. S. [1982]. *Syntactic Pattern Recognition and Applications*, Prentice Hall, Englewood Cliffs, N.J.
Fu, K. S. and Bhargava, B. K. [1973]. "Tree Systems for Syntactic Pattern Recognition," *IEEE Trans. Comput.*, vol. C-22, no. 12, pp. 1087–1099.
Fu, K. S., Gonzalez, R. C., and Lee, C. S. G. [1987]. *Robotics: Control, Sensing, Vision, and Intelligence*, McGraw-Hill, New York.
Fu, K. S. and Mui, J. K. [1981]. "A Survey of Image Segmentation," *Pattern Recog.*, vol. 13, no. 1, pp. 3–16.
Fukunaga, K. [1972]. *Introduction to Statistical Pattern Recognition*, Academic Press, New York.
Furht, B., Greenberg, J., and Westwater, R. [1997]. *Motion Estimation Algorithms for Video Compression*, Kluwer Academic Publishers, Boston.
Gallager, R. and Voorhis, D. V. [1975]. "Optimal Source Codes for Geometrically Distributed Integer Alphabets," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 228–230.
Gao, X., Sattar, F., and Vekateswarlu, R. [2007]. "Multiscale Corner Detection of Gray

Level Images Based on Log-Gabor Wavelet Transform," *IEEE Trans. Circuits and Systems for Video Technology*, in press.

Garcia, P. [1999]. "The Use of Boolean Model for Texture Analysis of Grey Images," *Computer Vision and Image Understanding*, vol. 74, no. 3, pp. 227–235.

Gdalyahu, Y. and Weinshall, D. [1999]. "Flexible Syntactic Matching of Curves and Its Application to Automated Hierarchical Classification of Silhouettes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 12, pp. 1312–1328.

Gegenfurtner, K. R. and Sharpe, L. T. (eds.) [1999]. *Color Vision: From Genes to Perception*, Cambridge University Press, New York.

Geladi, P. and Grahn, H. [1996]. *Multivariate Image Analysis*, John Wiley & Sons, New York.

Geman, D. and Reynolds, G. [1992]. "Constrained Restoration and the Recovery of Discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 3, pp. 367–383.

Gentleman, W. M. [1968]. "Matrix Multiplication and Fast Fourier Transformations," *Bell System Tech. J.*, vol. 47, pp. 1099–1103.

Gentleman, W. M. and Sande, G. [1966]. "Fast Fourier Transform for Fun and Profit," *Fall Joint Computer Conf.*, vol. 29, pp. 563–578, Spartan, Washington, D. C.

Gharavi, H. and Tabatabai, A. [1988]. "Sub-Band Coding of Monochrome and Color Images," *IEEE Trans. Circuits Sys.*, vol. 35, no. 2, pp. 207–214.

Giannakis, G. B. and Heath, R. W., Jr. [2000]. "Blind Identification of Multichannel FIR Blurs and Perfect Image Restoration," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1877–1896.

Giardina, C. R. and Dougherty, E. R. [1988]. *Morphological Methods in Image and Signal Processing*, Prentice Hall, Upper Saddle River, N.J.

Golomb, S. W. [1966]. "Run-Length Encodings," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 399–401.

Gonzalez, R. C. [1985]. "Computer Vision," in *Yearbook of Science and Technology*, McGraw-Hill, New York, pp. 128–132.

Gonzalez, R. C. [1985]. "Industrial Computer Vision," in *Advances in Information Systems Science*, Tou, J. T. (ed.), Plenum, New York, pp. 345–385.

Gonzalez, R. C. [1986]. "Image Enhancement and Restoration," in *Handbook of Pattern Recognition and Image Processing*, Young, T. Y., and Fu, K. S. (eds.), Academic Press, New York, pp. 191–213.

Gonzalez, R. C., Edwards, J. J., and Thomason, M. G. [1976]. "An Algorithm for the Inference of Tree Grammars," *Int. J. Comput. Info. Sci.*, vol. 5, no. 2, pp. 145–163.

Gonzalez, R. C. and Fittes, B. A. [1977]. "Gray-Level Transformations for Interactive Image Enhancement," *Mechanism and Machine Theory*, vol. 12, pp. 111–122.

Gonzalez, R. C. and Safabakhsh, R. [1982]. "Computer Vision Techniques for Industrial Applications," *Computer*, vol. 15, no. 12, pp. 17–32.

Gonzalez, R. C. and Thomason, M. G. [1978]. *Syntactic Pattern Recognition: An Introduction*, Addison-Wesley, Reading, Mass.

Gonzalez, R. C. and Woods, R. E. [1992]. *Digital Image Processing*, Addison-Wesley, Reading, Mass.

Gonzalez, R. C. and Woods, R. E. [2002]. *Digital Image Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. [2004]. *Digital Image Processing Using MATLAB*, Prentice Hall, Upper Saddle River, NJ.

Good, I. J. [1958]. "The Interaction Algorithm and Practical Fourier Analysis," *J. R. Stat. Soc. (Lond.)*, vol. B20, pp. 361–367; *Addendum*, vol. 22, 1960, pp. 372–375.

Goodson, K. J. and Lewis, P. H. [1990]. "A Knowledge-Based Line Recognition System," *Pattern Recog. Letters*, vol. 11, no. 4, pp. 295–304.

Gordon, I. E. [1997]. *Theories of Visual Perception*, 2nd ed., John Wiley & Sons, New York.

Gori, M. and Scarselli, F. [1998]. "Are Multilayer Perceptrons Adequate for Pattern Recognition and Verification?" *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 11, pp. 1121–1132.

Goutsias, J., Vincent, L., and Bloomberg, D. S. (eds) [2000]. *Mathematical Morphology and Its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Boston, Mass.
Graham, R. E. [1958]. "Predictive Quantizing of Television Signals," *IRE Wescon Conv. Rec.*, vol. 2, pt. 2, pp. 147–157.
Graham, R. L. and Yao, F. F. [1983]. "Finding the Convex Hull of a Simple Polygon," *J. Algorithms*, vol. 4, pp. 324–331.
Gray, R. M. [1984]. "Vector Quantization," *IEEE Trans. Acous. Speech Signal Processing*, vol. ASSP-1, no. 2, pp. 4–29.
Gröchenig, K. and Madych, W. R. [1992]. "Multiresolution Analysis, Haar Bases and Self-Similar Tilings of R^n," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 556–568.
Grossman, A. and Morlet, J. [1984]. "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape," *SIAM J. Appl. Math.* vol. 15, pp. 723–736.
Guil, N., Villalba, J. and Zapata, E. L. [1995]. "A Fast Hough Transform for Segment Detection," *IEEE Trans. Image Processing*, vol. 4, no. 11, pp. 1541–1548.
Guil, N. and Zapata, E. L. [1997]. "Lower Order Circle and Ellipse Hough Transform," *Pattern Recog.*, vol. 30, no. 10, pp. 1729–1744.
Gunn, S. R. [1998]. "Edge Detection Error in the Discrete Laplacian of a Gaussian," *Proc. 1998 Int'l Conference on Image Processing*, vol. II, pp. 515–519.
Gunn, S. R. [1999]. "On the Discrete Representation of the Laplacian of a Gaussian," *Pattern Recog.*, vol. 32, no. 8, pp. 1463–1472.
Gupta, L., Mohammad, R. S., and Tammana, R. [1990]. "A Neural Network Approach to Robust Shape Classification," *Pattern Recog.*, vol. 23, no. 6, pp. 563–568.
Gupta, L. and Srinath, M. D. [1988]. "Invariant Planar Shape Recognition Using Dynamic Alignment," *Pattern Recog.*, vol. 21, pp. 235–239.
Gupta, L., Wang., J., Charles, A., and Kisatsky, P. [1994]. "Three-Layer Perceptron Based Classifiers for the Partial Shape Classification Problem," *Pattern Recog.*, vol. 27, no. 1, pp. 91–97.
Haar, A. [1910]. "Zur Theorie der Orthogonalen Funktionensysteme," *Math. Annal.*, vol. 69, pp. 331–371.
Habibi, A. [1971]. "Comparison of *N*th Order DPCM Encoder with Linear Transformations and Block Quantization Techniques," *IEEE Trans. Comm. Tech.*, vol. COM-19, no. 6, pp. 948–956.
Habibi, A. [1974]. "Hybrid Coding of Pictorial Data," *IEEE Trans. Comm.*, vol. COM-22, no. 5, pp. 614–624.
Haddon, J. F. and Boyce, J. F. [1990]. "Image Segmentation by Unifying Region and Boundary Information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 10, pp. 929–948.
Hall, E. L. [1979]. *Computer Image Processing and Recognition*, Academic Press, New York.
Hamming, R. W. [1950]. "Error Detecting and Error Correcting Codes," *Bell Sys. Tech. J.*, vol. 29, pp. 147–160.
Hannah, I., Patel, D., and Davies, R. [1995]. "The Use of Variance and Entropy Thresholding Methods for Image Segmentation," *Pattern Recog.*, vol. 28, no. 8, pp. 1135–1143.
Haralick, R. M. and Lee, J. S. J. [1990]. "Context Dependent Edge Detection and Evaluation," *Pattern Recog.*, vol. 23, no. 1–2, pp. 1–20.
Haralick, R. M. and Shapiro, L. G. [1985]. "Survey: Image Segmentation," *Comput. Vision, Graphics, Image Proc.*, vol. 29, pp. 100–132.
Haralick, R. M. and Shapiro, L. G. [1992]. *Computer and Robot Vision*, vols. 1 & 2, Addison-Wesley, Reading, Mass.
Haralick, R. M., Sternberg, S. R., and Zhuang, X. [1987]. "Image Analysis Using Mathematical Morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 4, pp. 532–550.
Haralick, R. M., Shanmugan, R., and Dinstein, I. [1973]. "Textural Features for Image Classification," *IEEE Trans Syst. Man Cyb.*, vol. SMC-3, no. 6, pp. 610–621.
Harikumar, G. and Bresler, Y. [1999]. "Perfect Blind Restoration of Images Blurred by

Multiple Filters: Theory and Efficient Algorithms," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 202–219.
Harmuth, H. F. [1970]. *Transmission of Information by Orthogonal Signals*, Springer-Verlag, New York.
Haris, K., Efstratiadis, S. N, Maglaveras, N., and Katsaggelos, A. K. [1998]. "Hybrid Image Segmentation Using Watersheds and Fast Region Merging," *IEEE Trans. Image Processing*, vol. 7, no. 12, pp. 1684–1699.
Hart, P. E., Nilsson, N. J., and Raphael, B. [1968]. "A Formal Basis for the Heuristic Determination of Minimum-Cost Paths," *IEEE Trans. Syst. Man Cyb*, vol. SMC-4, pp. 100–107.
Hartenstein, H., Ruhl, M., and Saupe, D. [2000]. "Region-Based Fractal Image Compression," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1171–1184.
Haskell, B. G. and Netravali, A. N. [1997]. *Digital Pictures: Representation, Compression, and Standards*, Perseus Publishing, New York.
Haykin, S. [1996]. *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, N.J.
Healy, D. J. and Mitchell, O. R. [1981]. "Digital Video Bandwidth Compression Using Block Truncation Coding," *IEEE Trans. Comm.*, vol. COM-29, no. 12, pp. 1809–1817.
Heath, M. D., Sarkar, S., Sanocki, T., and Bowyer, K. W. [1997]. "A Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 12, pp. 1338–1359.
Heath, M., Sarkar, S., Sanoki, T., and Bowyer, K. [1998]. "Comparison of Edge Detectors: A Methodology and Initial Study," *Computer Vision and Image Understanding*, vol. 69, no. 1, pp. 38–54.
Hebb, D. O. [1949]. *The Organization of Behavior: A Neuropsychological Theory*, John Wiley & Sons, New York.
Heijmans, H. J. A. M. and Goutsias, J. [2000]. "Nonlinear Multiresolution Signal Decomposition Schemes—Part II: Morphological Wavelets," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1897–1913.
Highnam, R. and Brady, M. [1997]. "Model-Based Image Enhancement of Far Infrared Images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 4, pp. 410–415.
Hojjatoleslami, S. A. and Kittler, J. [1998]. "Region Growing: A New Approach," *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1079–1084.
Hong, Pi, Hung, Li, and Hua, Li [2006]. "A Novel Fractal Image Watermarking," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 488–499.
Hoover, R. B. and Doty, F. [1996]. *Hard X-Ray/Gamma-Ray and Neutron Optics, Sensors, and Applications*, SPIE Press, Bellingham, Wash.
Horn, B. K. P. [1986]. *Robot Vision*, McGraw-Hill, New York.
Hotelling, H. [1933]. "Analysis of a Complex of Statistical Variables into Principal Components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, 498–520.
Hough, P. V. C. [1962]. "Methods and Means for Recognizing Complex Patterns," U. S. Patent 3,069,654.
Hsu, C. C. and Huang, J. S. [1990]. "Partitioned Hough Transform for Ellipsoid Detection," *Pattern Recog.*, vol. 23, no. 3–4, pp. 275–282.
Hu. J. and Yan, H. [1997]. "Polygonal Approximation of Digital Curves Based on the Principles of Perceptual Organization," *Pattern Recog.*, vol. 30, no. 5, pp. 701–718.
Hu, M. K. [1962]. "Visual Pattern Recognition by Moment Invariants," *IRE Trans. Info. Theory*, vol. IT-8, pp. 179–187.
Hu, Y., Kwong, S., and Huang, J. [2006]. "An Algorithm for Removable Visible Watermarking," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 129–133.
Huang, K.-Y. [2002]. *Syntactic Pattern Recognition for Seismic Oil Exploration*, World Scientific, Hackensack, NJ.
Huang, S.-C. and Sun, Y.-N. [1999]. "Polygonal Approximation Using Generic Algorithms," *Pattern Recog.*, vol. 32, no. 8, pp. 1409–1420.
Huang, T. S. [1965]. "PCM Picture Transmission," *IEEE Spectrum*, vol. 2, no. 12, pp. 57–63.
Huang, T. S. [1966]. "Digital Picture Coding," *Proc. Natl. Electron. Conf.*, pp. 793–797.
Huang, T. S., ed. [1975]. *Picture Processing and Digital Filtering*, Springer-Verlag,

New York.
Huang, T. S. [1981]. *Image Sequence Analysis*, Springer-Verlag, New York.
Huang, T. S. and Hussian, A. B. S. [1972]. "Facsimile Coding by Skipping White," *IEEE Trans. Comm.*, vol. COM-23, no. 12, pp. 1452–1466.
Huang, T. S. and Tretiak, O. J. (eds.). [1972]. *Picture Bandwidth Compression*, Gordon and Breech, New York.
Huang, T. S., Yang, G. T., and Tang, G. Y. [1979]. "A Fast Two-Dimensional Median Filtering Algorithm," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-27, pp. 13–18.
Huang, Y. and Schultheiss, P. M. [1963]. "Block Quantization of Correlated Gaussian Random Variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 289–296.
Hubbard, B. B. [1998]. *The World According to Wavelets—The Story of a Mathematical Technique in the Making*, 2nd ed, A. K. Peters, Ltd., Wellesley, Mass.
Hubel, D. H. [1988]. *Eye, Brain, and Vision*, Scientific Amer. Library, W. H. Freeman, New York.
Huertas, A. and Medione, G. [1986]. "Detection of Intensity Changes with Subpixel Accuracy using Laplacian-Gaussian Masks," *IEEE Trans. Pattern. Anal. Machine Intell.*, vol. PAMI-8, no. 5, pp. 651–664.
Huffman, D. A. [1952]. "A Method for the Construction of Minimum Redundancy Codes," *Proc. IRE*, vol. 40, no. 10, pp. 1098–1101.
Hufnagel, R. E. and Stanley, N. R. [1964]. "Modulation Transfer Function Associated with Image Transmission Through Turbulent Media," *J. Opt. Soc. Amer.*, vol. 54, pp. 52–61.
Hummel, R. A. [1974]. "Histogram Modification Techniques," Technical Report TR-329. F-44620–72C-0062, Computer Science Center, University of Maryland, College Park, Md.
Hunt, B. R. [1971]. "A Matrix Theory Proof of the Discrete Convolution Theorem," *IEEE Trans. Audio and Electroacoust.*, vol. AU-19, no. 4, pp. 285–288.
Hunt, B. R. [1973]. "The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer," *IEEE Trans. Comput.*, vol. C-22, no. 9, pp. 805–812.
Hunter, R. and Robinson, A. H. [1980]. "International Digital Facsimile Coding Standards," *Proc. IEEE*, vol. 68, no. 7, pp. 854–867.
Hurn, M. and Jennison, C. [1996]. "An Extension of Geman and Reynolds' Approach to Constrained Restoration and the Recovery of Discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 6, pp. 657–662.
Hwang, H. and Haddad, R. A. [1995]. "Adaptive Median Filters: New Algorithms and Results," *IEEE Trans. Image Processing*, vol. 4, no. 4, pp. 499–502.
IEEE Computer [1974]. Special issue on digital image processing. vol. 7; no. 5.
IEEE Computer [1988]. Special issue on artificial neural systems. vol. 21, no. 3.
IEEE Trans. Circuits and Syst. [1975]. Special issue on digital filtering and image processing, vol. CAS-2, pp. 161–304.
IEEE Trans. Computers [1972]. Special issue on two-dimensional signal processing, vol. C-21, no. 7.
IEEE Trans. Comm. [1981]. Special issue on picture communication systems, vol. COM-29, no. 12.
IEEE Trans. on Image Processing [1994]. Special issue on image sequence compression, vol. 3, no. 5.
IEEE Trans. on Image Processing [1996]. Special issue on vector quantization, vol. 5, no. 2.
IEEE Trans. Image Processing [1996]. Special issue on nonlinear image processing, vol. 5, no. 6.
IEEE Trans. Image Processing [1997]. Special issue on automatic target detection, vol. 6, no. 1.
IEEE Trans. Image Processing [1997]. Special issue on color imaging, vol. 6, no. 7.
IEEE Trans. Image Processing [1998]. Special issue on applications of neural networks to image processing, vol. 7, no. 8.
IEEE Trans. Information Theory [1992]. Special issue on wavelet transforms and

mulitresolution signal analysis, vol. 11, no. 2, Part II.
IEEE Trans. Pattern Analysis and Machine Intelligence [1989]. Special issue on multiresolution processing, vol. 11, no. 7.
IEEE Trans. Signal Processing [1993]. Special issue on wavelets and signal processing, vol. 41, no. 12.
IES Lighting Handbook, 9th ed. [2000]. Illuminating Engineering Society Press, New York.
ISO/IEC [1999]. *ISO/IEC 14495-1:1999: Information technology—Lossless and near-lossless compression of continuous-tone still images: Baseline.*
ISO/IEC JTC 1/SC 29/WG 1 [2000]. *ISO/IEC FCD 15444-1: Information technology—JPEG 2000 image coding system: Core coding system.*
Jähne, B. [1997]. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*, Springer-Verlag, New York.
Jähne, B. [2002]. *Digital Image Processing*, 5th ed., Springer, New York.
Jain, A. K. [1981]. "Image Data Compression: A Review," *Proc. IEEE*, vol. 69, pp. 349–389.
Jain, A. K. [1989]. *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, N.J.
Jain, A. K., Duin, R. P. W., and Mao, J. [2000]. "Statistical Pattern Recognition: A Review," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 1, pp. 4–37.
Jain, J. R. and Jain, A. K. [1981]. "Displacement Measurement and Its Application in Interframe Image Coding," *IEEE Trans. Comm.*, vol. COM-29, pp. 1799–1808.
Jain, R. [1981]. "Dynamic Scene Analysis Using Pixel-Based Processes," *Computer*, vol. 14, no. 8, pp. 12–18.
Jain, R., Kasturi, R., and Schunk, B. [1995]. *Computer Vision*, McGraw-Hill, New York.
Jang, B. K. and Chin, R. T. [1990]. "Analysis of Thinning Algorithms Using Mathematical Morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 6, pp. 541–551.
Jayant, N. S. (ed.) [1976]. *Waveform Quantization and Coding*, IEEE Press, New York.
Jones, R. and Svalbe, I. [1994]. "Algorithms for the Decomposition of Gray-Scale Morphological Operations,"*IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 6, pp. 581–588.
Jonk, A., van den Boomgaard, S., and Smeulders, A. [1999]. "Grammatical Inference of Dashed Lines," *Computer Vision and Image Understanding*, vol. 74, no. 3, pp. 212–226.
Kahaner, D. K. [1970]. "Matrix Description of the Fast Fourier Transform," *IEEE Trans. Audio Electroacoustics*, vol. AU-18, no. 4, pp. 442–450.
Kak, A. C. and Slaney, M. [2001]. *Principles of Computerized Tomographic Imaging*, Society for Industrial and Applied Mathematics, Philadelphia, Pa.
Kamstra, L. and Heijmans, H.J.A.M. [2005]. "Reversible Data Embedding Into Images Using Wavelet Techniques and Sorting," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2082–2090.
Karhunen, K. [1947]. "Über Lineare Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fennicae*, Ser. A137. (Translated by I. Selin in "On Linear Methods in Probability Theory." T-131, 1960, The RAND Corp., Santa Monica, Calif.)
Kasson, J. and Plouffe, W. [1992]. "An Analysis of Selected Computer Interchange Color Spaces," *ACM Trans. on Graphics*, vol. 11, no. 4, pp. 373–405.
Katzir, N., Lindenbaum, M., and Porat, M. [1994]. "Curve Segmentation Under Partial Occlusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 5, pp. 513–519.
Kerre, E. E. and Nachtegael, M., eds. [2000]. *Fuzzy Techniques in Image Processing*, Springer-Verlag, New York.
Khanna, T. [1990]. *Foundations of Neural Networks*, Addison-Wesley, Reading, Mass.
Kim, C. [2005]. "Segmenting a Low-Depth-of-Filed Image Using Morphological Filters and Region Merging," *IEEE Trans. Image Proc.*, vol. 14, no. 10, pp. 1503–1511.
Kim, J. K., Park, J. M., Song, K. S., and Park, H. W. [1997]. "Adaptive Mammographic Image Enhancement Using First Derivative and Local Statistics," *IEEE Trans. Medical Imaging*, vol. 16, no. 5, pp. 495–502.
Kimme, C., Ballard, D. H., and Sklansky, J. [1975]. "Finding Circles by an Array of Accumulators," *Comm. ACM*, vol. 18, no. 2, pp. 120–122.
Kirsch, R. [1971]. "Computer Determination of the Constituent Structure of Biological

Images," *Comput. Biomed. Res.*, vol. 4, pp. 315–328.
Kiver, M. S. [1965]. *Color Television Fundamentals*, McGraw-Hill, New York.
Klette, R. and Rosenfeld, A. [2004]. *Digital Geometry—Geometric Methods for Digital Picture Analysis*, Morgan Kaufmann, San Francisco.
Klinger, A. [1976]. "Experiments in Picture Representation Using Regular Decomposition," *Comput. Graphics Image Proc.*, vol. 5, pp. 68–105.
Knowlton, K. [1980]. "Progressive Transmission of Gray-Scale and Binary Pictures by Simple, Efficient, and Lossless Encoding Schemes," *Proc. IEEE*, vol. 68, no. 7, pp. 885–896.
Kohler, R. J. and Howell, H. K. [1963]. "Photographic Image Enhancement by Superposition of Multiple Images," *Photogr. Sci. Eng.*, vol. 7, no. 4, pp. 241–245.
Kokaram, A. [1998]. *Motion Picture Restoration*, Springer-Verlag, New York.
Kokare, M., Biswas, P., and Chatterji, B. [2005]. "Texture Image Retrieval Using New Rotated Complex Wavelet Filters," *IEEE Trans. Systems, Man, Cybernetics, Part B*, vol. 35, no. 6, pp. 1168–1178.
Kramer, H. P. and Mathews, M. V. [1956]. "A Linear Coding for Transmitting a Set of Correlated Signals," *IRE Trans. Info. Theory*, vol. IT-2, pp. 41–46.
Langdon, G. C. and Rissanen, J. J. [1981]. "Compression of Black-White Images with Arithmetic Coding," *IEEE Trans. Comm.*, vol. COM-29, no. 6, pp. 858–867.
Lantuéjoul, C. [1980]. "Skeletonization in Quantitative Metallography," in *Issues of Digital Image Processing*, Haralick, R. M., and Simon, J. C. (eds.), Sijthoff and Noordhoff, Groningen, The Netherlands.
Latecki, L. J. and Lakämper, R. [1999]. "Convexity Rule for Shape Decomposition Based on Discrete Contour Evolution," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 441–454.
Le Gall, D. and Tabatabai, A. [1988]. "Sub-Band Coding of Digital Images Using Symmetric Short Kernel Filters and Arithmetic Coding Techniques," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, pp. 761–765.
Ledley, R. S. [1964]. "High-Speed Automatic Analysis of Biomedical Pictures," *Science*, vol. 146, no. 3461, pp. 216–223.
Lee, J.-S., Sun, Y.-N., and Chen, C.-H. [1995]. "Multiscale Corner Detection by Using Wavelet Transforms," *IEEE Trans. Image Processing*, vol. 4, no. 1, pp. 100–104.
Lee, S. U., Chung, S. Y., and Park, R. H. [1990]. "A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation," *Comput. Vision, Graphics, Image Proc.*, vol. 52, no. 2, pp. 171–190.
Lehmann, T. M., Gönner, C., and Spitzer, K. [1999]. "Survey: Interpolation Methods in Medical Image Processing," *IEEE Trans. Medical Imaging*, vol. 18, no. 11, pp. 1049–1076.
Lema, M. D. and Mitchell, O. R. [1984]. "Absolute Moment Block Truncation Coding and Its Application to Color Images," *IEEE Trans. Comm.*, vol. COM-32, no. 10. pp. 1148–1157.
Levine, M. D. [1985]. *Vision in Man and Machine*, McGraw-Hill, New York.
Liang, K.-C. and Kuo, C.-C. J. [1991]. "Waveguide: A Joint Wavelet-Based Image Representation and Description System," *IEEE Trans. Image Processing*, vol. 8, no. 11, pp. 1619–1629.
Liang, Q., Wendelhag, J. W., and Gustavsson, T. [2000]. "A Multiscale Dynamic Programming Procedure for Boundary Detection in Ultrasonic Artery Images," *IEEE Trans. Medical Imaging*, vol. 19, no. 2, pp. 127–142.
Liao, P., Chen, T., and Chung, P. [2001]. "A Fast Algorithm for Multilevel Thresholding," *J. Inform. Sc. and Eng.*, vol. 17, pp. 713–727.
Lillesand, T. M. and Kiefer, R. W. [1999]. *Remote Sensing and Image Interpretation*, John Wiley & Sons, New York.
Lim, J. S. [1990]. *Two-Dimensional Signal and Image Processing*, Prentice Hall, Upper Saddle River, N.J.
Limb, J. O. and Rubinstein, C. B. [1978]. "On the Design of Quantizers for DPCM Coders: A Functional Relationship Between Visibility, Probability, and Masking,"

IEEE Trans. Comm., vol. COM-26, pp. 573–578.
Lindblad, T. and Kinser, J. M. [1998]. *Image Processing Using Pulse-Coupled Neural Networks*, Springer-Verlag, New York.
Linde, Y., Buzo, A., and Gray, R. M. [1980]. "An Algorithm for Vector Quantizer Design," *IEEE Trans. Comm.*, vol. COM-28, no. 1, pp. 84–95.
Lippmann, R. P. [1987]. "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, vol. 4, pp. 4–22.
Liu, J. and Yang, Y.-H. [1994]. "Multiresolution Color Image Segmentation," *IEEE Trans Pattern Anal. Machine Intell.*, vol. 16, no. 7, pp. 689–700.
Liu-Yu, S. and Antipolis, M. [1993]. "Description of Object Shapes by Apparent Boundary and Convex Hull," *Pattern Recog.*, vol. 26, no. 1, pp. 95–107.
Lo, R.-C. and Tsai, W.-H. [1995]. "Gray-Scale Hough Transform for Thick Line Detection in Gray-Scale Images," *Pattern Recog.*, vol. 28, no. 5, pp. 647–661.
Loncaric, S. [1998]. "A Survey of Shape Analysis Techniques," *Pattern Recog.*, vol. 31, no. 8, pp. 983–1010.
Lu, H. E. and Wang, P. S. P. [1986]. "A Comment on 'A Fast Parallel Algorithm for Thinning Digital Patterns,'" *Comm. ACM*, vol. 29, no. 3, pp. 239–242.
Lu, N. [1997]. *Fractal Imaging*, Academic Press, New York.
Lu, W.-S. and Antoniou, A. [1992]. "Two-Dimensional Digital Filters," Marcel Dekker, New York.
MacAdam, D. L. [1942]. "Visual Sensitivities to Color Differences in Daylight," *J. Opt. Soc. Am.*, vol. 32, pp. 247–274.
MacAdam, D. P. [1970]. "Digital Image Restoration by Constrained Deconvolution," *J. Opt. Soc. Am.*, vol. 60, pp. 1617–1627.
Maki, A., Nordlund, P., and Eklundh, J.-O. [2000]. "Attentional Scene Segmentation: Integrating Depth and Motion," *Computer Vision and Image Understanding*, vol. 78, no. 3, pp. 351–373.
Malacara, D. [2001]. *Color Vision and Colorimetry: Theory and Applications*, SPIE Press, Bellingham, Wash.
Mallat, S. [1987]. "A Compact Multiresolution Representation: The Wavelet Model," *Proc. IEEE Computer Society Workshop on Computer Vision*, IEEE Computer Society Press, Washington, D. C., pp. 2–7.
Mallat, S. [1989a]. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, pp. 674–693.
Mallat, S. [1989b]. "Multiresolution Approximation and Wavelet Orthonormal Bases of L^2," *Trans. American Mathematical Society*, vol. 315, pp. 69–87.
Mallat, S. [1989c]. "Multifrequency Channel Decomposition of Images and Wavelet Models," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, pp. 2091–2110.
Mallat, S. [1998]. *A Wavelet Tour of Signal Processing*, Academic Press, Boston, Mass.
Mallat, S. [1999]. *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, San Diego, Calif.
Mallot, A. H. [2000]. *Computational Vision*, The MIT Press, Cambridge, Mass.
Mamistvalov, A. [1998]. "*n*-Dimensional Moment Invariants and Conceptual Mathematical Theory of Recognition [of] *n*-Dimensional Solids," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 8, pp. 819–831.
Manjunath, B., Salembier, P., and Sikora, T. [2001]. *Introduction to MPEG-7*, John Wiley & Sons, West Sussex, UK.
Maragos, P. [1987]. "Tutorial on Advances in Morphological Image Processing and Analysis," Optical Engineering, vol. 26, no. 7, pp. 623–632.
Marchand-Maillet, S. and Sharaiha, Y. M. [2000]. *Binary Digital Image Processing: A Discrete Approach*, Academic Press, New York.
Maren, A. J., Harston, C. T., and Pap, R. M. [1990]. *Handbook of Neural Computing Applications*, Academic Press, New York.
Marr, D. [1982]. *Vision*, Freeman, San Francisco.

Marr, D. and Hildreth, E. [1980]. "Theory of Edge Detection," *Proc. R. Soc. Lond.*, vol. B207, pp. 187–217.

Martelli, A. [1972]. "Edge Detection Using Heuristic Search Methods," *Comput. Graphics Image Proc.*, vol. 1, pp. 169–182.

Martelli, A. [1976]. "An Application of Heuristic Search Methods to Edge and Contour Detection," *Comm. ACM*, vol. 19, no. 2, pp. 73–83.

Martin, M. B. and Bell, A. E. [2001]. "New Image Compression Techniques Using Multiwavelets and Multiwavelet Packets," *IEEE Trans. on Image Proc.*, vol. 10, no. 4, pp. 500–510.

Mather, P. M. [1999]. *Computer Processing of Remotely Sensed Images: An Introduction*, John Wiley & Sons, New York.

Max, J. [1960]. "Quantizing for Minimum Distortion," *IRE Trans. Info. Theory*, vol. IT-6, pp. 7–12.

McClelland, J. L. and Rumelhart, D. E. (eds.) [1986]. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 2: *Psychological and Biological Models*, The MIT Press, Cambridge, Mass.

McCulloch, W. S. and Pitts, W. H. [1943]. "A Logical Calculus of the Ideas Imminent in Nervous Activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133.

McFarlane, M. D. [1972]. "Digital Pictures Fifty Years Ago," *Proc. IEEE*, vol. 60, no. 7, pp. 768–770.

McGlamery, B. L. [1967]. "Restoration of Turbulence-Degraded Images," *J. Opt. Soc. Am.*, vol. 57, no. 3, pp. 293–297.

Meijering, H. W., Zuiderveld, K. J., and Viergever, M. A. [1999]. "Image Registration for Digital Subtraction Angiography," *Int. J. Comput. Vision*, vol. 31, pp. 227–246.

Meijering, E. H. W., Niessen, W. J., and Viergever, M. A. [1999]. "Retrospective Motion Correction in Digital Subtraction Angiography: A Review," *IEEE Trans. Medical Imaging*, vol. 18, no. 1, pp. 2–21.

Meijering, E. H. W., et al. [2001]. "Reduction of Patient Motion Artifacts in Digital Subtraction Angiography: Evaluation of a Fast and Fully Automatic Technique," *Radiology*, vol. 219, pp. 288–293.

Memon, N., Neuhoff, D. L., and Shende, S. [2000]. "An Analysis of Some Common Scanning Techniques for Lossless Image Coding," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1837–1848.

Mesarović, V. Z. [2000]. "Iterative Linear Minimum Mean-Square-Error Image Restoration from Partially Known Blur," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 17, no. 4, pp. 711–723.

Meyer, Y. [1987]. "L'analyses par Ondelettes," *Pour la Science*.

Meyer, Y. [1990]. *Ondelettes et ope'rateurs*, Hermann, Paris.

Meyer, Y. (ed.) [1992a]. *Wavelets and Applications: Proceedings of the International Conference, Marseille, France*, Mason, Paris, and Springer-Verlag, Berlin.

Meyer, Y. (translated by D. H. Salinger) [1992b]. *Wavelets and Operators*, Cambridge University Press, Cambridge, UK.

Meyer, Y. (translated by R. D. Ryan) [1993]. *Wavelets: Algorithms and Applications*, Society for Industrial and Applied Mathematics, Philadelphia.

Meyer, F. G., Averbuch, A. Z., and Strömberg, J.-O. [2000]. "Fast Adaptive Wavelet Packet Image Compression," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 792–800.

Meyer, F. and Beucher, S. [1990]. "Morphological Segmentation," *J. Visual Comm., and Image Representation*, vol. 1, no. 1, pp. 21–46.

Meyer, H., Rosdolsky, H. G., and Huang, T. S. [1973]. "Optimum Run Length Codes," *IEEE Trans. Comm.*, vol. COM-22, no. 6, pp. 826–835.

Minsky, M. and Papert, S. [1969]. *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass.

Mirmehdi, M. and Petrou, M. [2000]. "Segmentation of Color Textures," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 2, pp. 142–159.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J.-M. [1996]. *Wavelet Toolbox User's Guide*, The MathWorks, Inc., Natick, Mass.

Mitchell, D. P. and Netravali, A. N. [1988]. "Reconstruction Filters in Computer Graphics," *Comp. Graphics*, vol. 22, no. 4, pp. 221–228.
Mitchell, J., Pennebaker, W., Fogg, C., and LeGall, D. [1997]. *MPEG Video Compression Standard*, Chapman & Hall, New York.
Mitiche, A. [1994]. *Computational Analysis of Visual Motion*, Perseus Publishing, New York.
Mitra, S., Murthy, C., and Kundu, M. [1998]. "Technique for Fractal Image Compression Using Genetic Algorithm," *IEEE Transactions on Image Processing*, vol. 7, no. 4, pp. 586–593.
Mitra, S. K. and Sicuranza, G. L. (eds.) [2000]. *Nonlinear Image Processing*, Academic Press, New York.
Mohanty, S., et al. [1999]. "A Dual Watermarking Technique for Images," *Proc.* 7th *ACM International Multimedia Conference*, ACM-MM'99, Part 2, pp. 49–51.
Moore, G. A. [1968]. "Automatic Scanning and Computer Processes for the Quantitative Analysis of Micrographs and Equivalent Subjects," in *Pictorial Pattern Recognition*, (G. C. Cheng et al., eds), pp. 275–326, Thomson, Washington, D.C.
Mukherjee, D. and Mitra, S. [2003]. "Vector SPIHT for Embedded Wavelet Video and Image Coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 3, pp. 231–246.
Murase, H. and Nayar, S. K. [1994]. "Illumination Planning for Object Recognition Using Parametric Eigen Spaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 12, pp. 1219–1227.
Murino, V., Ottonello, C., and Pagnan, S. [1998]. "Noisy Texture Classfication: A Higher-Order Statistical Approach," *Pattern Recog.*, vol. 31, no. 4, pp. 383–393.
Nagao, M. and Matsuyama, T. [1980]. *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York.
Najman, L. and Schmitt, M. [1996]. "Geodesic Saliency of Watershed Contours and Hierarchical Segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 12, pp. 1163–1173.
Narendra, P. M. and Fitch, R. C. [1981]. "Real-Time Adaptive Contrast Enhancement," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, no. 6, pp. 655–661.
Netravali, A. N. [1977]. "On Quantizers for DPCM Coding of Picture Signals," *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, pp. 360–370.
Netravali, A. N. and Limb, J. O. [1980]. "Picture Coding: A Review," *Proc. IEEE*, vol. 68, no. 3, pp. 366–406.
Nevatia, R. [1982]. *Machine Perception*, Prentice Hall, Upper Saddle River, N.J.
Ngan, K. N., Meier, T., and Chai, D. [1999]. *Advanced Video Coding: Principles and Techniques*, Elsevier, Boston.
Nie, Y. and Barner, K. E. [2006]. "The Fuzzy Transformation and Its Applications in Image Processing," IEEE Trans. Image Proc., vol. 15, no. 4, pp. 910–927.
Nilsson, N. J. [1965]. *Learning Machines: Foundations of Trainable Pattern-Classifying Systems*, McGraw-Hill, New York.
Nilsson, N. J. [1971]. *Problem Solving Methods in Artificial Intelligence*, McGraw-Hill, New York.
Nilsson, N. J. [1980]. *Principles of Artificial Intelligence*, Tioga, Palo Alto, Calif.
Nixon, M. and Aguado, A. [2002]. *Feature Extraction and Image Processing*, Newnes, Boston, MA.
Noble, B. and Daniel, J. W. [1988]. *Applied Linear Algebra*, 3rd ed., Prentice Hall, Upper Saddle River, N.J.
O'Connor, Y. Z. and Fessler, J. A. [2006]. "Fourier-Based Forward and Back-Projections in Iterative Fan-Beam Tomographic Image Reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, no. 5, pp. 582–589.
Odegard, J. E., Gopinath, R. A., and Burrus, C. S. [1992]. "Optimal Wavelets for Signal Decomposition and the Existence of Scale-Limited Signals," *Proceedings of IEEE Int. Conf. On Signal Proc.*, ICASSP-92, San Francisco, CA, vol. IV, 597–600.
Olkkonen, J. and Olkkonen, H. [2007]. Discrete Lattice Wavelet Transform, *IEEE*

Trans. Circuits and Systems II: Express Briefs, vol. 54, no. 1, pp. 71–75.
Olson, C. F. [1999]. "Constrained Hough Transforms for Curve Detection," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 329–345.
O'Neil, J. B. [1971]. "Entropy Coding in Speech and Television Differential PCM Systems," *IEEE Trans. Info. Theory*, vol. IT-17, pp. 758–761.
Oommen, R. J. and Loke, R. K. S. [1997]. "Pattern Recognition of Strings with Substitutions, Insertions, Deletions, and Generalized Transpositions," *Pattern Recog.*, vol. 30, no. 5, pp. 789–800.
Oppenheim, A. V. and Schafer, R. W. [1975]. *Digital Signal Processing*, Prentice Hall, Englewood Cliffs, N.J.
Oppenheim, A. V., Schafer, R. W., and Stockham, T. G., Jr. [1968]. "Nonlinear Filtering of Multiplied and Convolved Signals," *Proc. IEEE*, vol. 56, no. 8, pp. 1264–1291.
Oster, G. and Nishijima, Y. [1963]. "Moiré Patterns," *Scientific American*, vol. 208, no. 5, pp. 54–63.
Otsu, N. [1979]. "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66.
Oyster, C. W. [1999]. *The Human Eye: Structure and Function*, Sinauer Associates, Sunderland, Mass.
Paez, M. D. and Glisson, T. H. [1972]. "Minimum Mean-Square-Error Quantization in Speech PCM and DPCM Systems," *IEEE Trans. Comm.*, vol. COM-20, pp. 225–230.
Pao, Y. H. [1989]. *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Reading, Mass.
Papamarkos, N. and Atsalakis, A. [2000]. "Gray-Level Reduction Using Local Spatial Features," *Computer Vision and Image Understanding*, vol. 78, no. 3, pp. 336–350.
Papoulis, A. [1991]. *Probability, Random Variables, and Stochastic Processes*, 3rd ed., McGraw-Hill, New York.
Parhi, K. and Nishitani, T. [1999]. "Digital Signal Processing in Multimedia Systems," Chapter 18: *A Review of Watermarking Principles and Practices*, M. Miller, et al., pp. 461–485, Marcel Dekker Inc., New York.
Park, H. and Chin, R. T. [1995]. "Decomposition of Arbitrarily-Shaped Morphological Structuring," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 1, pp. 2–15.
Parker, J. R. [1991]. "Gray Level Thresholding in Badly Illuminated Images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 8, pp. 813–819.
Parker, J. R. [1997]. *Algorithms for Image Processing and Computer Vision*, John Wiley & Sons, New York.
Patrascu, V. [2004]. "Fuzzy Enhancement Method Using Logarithmic Model," *IEEE-Fuzz'04*, vol. 3, pp. 1431–1436.
Pattern Recognition [2000]. Special issue on mathematical morphology and nonlinear image processing, vol. 33, no. 6, pp. 875–1117.
Pavlidis, T. [1977]. *Structural Pattern Recognition*, Springer-Verlag, New York.
Pavlidis, T. [1982]. *Algorithms for Graphics and Image Processing*, Computer Science Press, Rockville, Md.
Pavlidis, T. and Liow, Y. T. [1990]. "Integrating Region Growing and Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 3, pp. 225–233.
Peebles, P. Z. [1993]. *Probability, Random Variables, and Random Signal Principles*, 3rd ed., McGraw-Hill, New York.
Pennebaker, W. B. and Mitchell, J. L. [1992]. *JPEG: Still Image Data Compression Standard*, Van Nostrand Reinhold, New York.
Pennebaker, W. B., Mitchell, J. L., Langdon, G. G., Jr., and Arps, R. B. [1988]. "An Overview of the Basic Principles of the Q-coder Adaptive Binary Arithmetic Coder," *IBM J. Res. Dev.*, vol. 32, no. 6, pp. 717–726.
Perez, A. and Gonzalez, R. C. [1987]. "An Iterative Thresholding Algorithm for Image Segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 6, pp. 742–751.
Perona, P. and Malik, J. [1990]. "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 7, pp. 629–639.
Persoon, E. and Fu, K. S. [1977]. "Shape Discrimination Using Fourier Descriptors,"

IEEE Trans. Systems Man Cyb., vol. SMC-7, no. 2, pp. 170–179.
Petrou, M. and Bosdogianni, P. [1999]. *Image Processing: The Fundamentals*, John Wiley & Sons, UK.
Petrou, M. and Kittler, J. [1991]. "Optimal Edge Detector for Ramp Edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 5, pp. 483–491.
Piech, M. A. [1990]. "Decomposing the Laplacian," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 8, pp. 830–831.
Pitas, I. and Vanetsanopoulos, A. N. [1990]. *Nonlinear Digital Filters: Principles and Applications*, Kluwer Academic Publishers, Boston, Mass.
Plataniotis, K. N. and Venetsanopoulos, A. N. [2000]. *Color Image Processing and Applications*, Springer-Verlag, New York.
Pokorny, C. K. and Gerald, C. F. [1989]. *Computer Graphics: The Principles Behind the Art and Science*, Franklin, Beedle & Associates, Irvine, Calif.
Porco, C. C., West R. A., et al. [2004]. "Cassini Imaging Science: Instrument Characteristics and Anticipated Scientific Investigations at Saturn," *Space Science Reviews*, vol. 115, pp. 363–497.
Poynton, C. A. [1996]. *A Technical Introduction to Digital Video*, John Wiley & Sons, New York.
Prasad, L. and Iyengar, S. S. [1997]. *Wavelet Analysis with Applications to Image Processing*, CRC Press, Boca Raton, Fla.
Pratt, W. K. [2001]. *Digital Image Processing*, 3rd ed., John Wiley & Sons, New York.
Preparata, F. P. and Shamos, M. I. [1985]. *Computational Geometry: An Introduction*, Springer-Verlag, New York.
Preston, K. [1983]. "Cellular Logic Computers for Pattern Recognition," *Computer*, vol. 16, no. 1, pp. 36–47.
Prewitt, J. M. S. [1970]. "Object Enhancement and Extraction," in *Picture Processing and Psychopictorics*, Lipkin, B. S., and Rosenfeld, A. (eds.), Academic Press, New York.
Prince, J. L. and Links, J. M. [2006]. *Medical Imaging Signals and Systems*, Prentice Hall, Upper Saddle River, NJ.
Principe, J. C., Euliano, N. R., and Lefebre, W. C. [1999]. *Neural and Adaptive Systems: Fundamentals through Simulations*, John Wiley & Sons, New York.
Pritchard, D. H. [1977]. "U. S. Color Television Fundamentals—A Review," *IEEE Trans. Consumer Electronics*, vol. CE-23, no. 4, pp. 467–478.
Proc. IEEE [1967]. Special issue on redundancy reduction, vol. 55, no. 3.
Proc. IEEE [1972]. Special issue on digital picture processing, vol. 60, no. 7.
Proc. IEEE [1980]. Special issue on the encoding of graphics, vol. 68, no. 7.
Proc. IEEE [1985]. Special issue on visual communication systems, vol. 73, no. 2.
Qian, R. J. and Huang, T. S. [1996]. "Optimal Edge Detection in Two-Dimensional Images," *IEEE Trans. Image Processing*, vol. 5, no. 7, pp. 1215–1220.
Rabbani, M. and Jones, P. W. [1991]. *Digital Image Compression Techniques*, SPIE Press, Bellingham, Wash.
Rajala, S. A., Riddle, A. N., and Snyder, W. E. [1983]. "Application of One-Dimensional Fourier Transform for Tracking Moving Objects in Noisy Environments," *Comp., Vision, Image Proc.*, vol. 21, pp. 280–293.
Ramachandran, G. N. and Lakshminarayanan, A. V. [1971]. "Three Dimensional Reconstructions from Radiographs and Electron Micrographs: Application of Convolution Instead of Fourier Transforms," Proc. Nat. Acad. Sci., vol. 68, pp. 2236–2240.
Rane, S. and Sapiro, G. [2001]. "Evaluation of JPEG-LS, the New Lossless and Controlled-Lossy Still Image Compression Standard, for Compression of High-Resolution Elevation Data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, no. 10, pp. 2298–2306.
Rangayyan, R. M. [2005]. *Biomedical Image Analysis*, CRC Press, Boca Raton, FL.
Reddy, B. S. and Chatterji, B. N. [1996]. "An FFT-Based Technique for Translation, Rotation, and Scale Invariant Image Registration," *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1266–1271.
Regan, D. D. [2000]. *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion, and Binocular Disparity*, Sin-

auer Associates, Sunderland, Mass.

Rice, R. F. [1979]. "Some Practical Universal Noiseless Coding Techniques," *Tech. Rep. JPL-79-22*, Jet Propulsion Lab., Pasadena, CA.

Ritter, G. X. and Wilson, J. N. [2001]. *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, Boca Raton, Fla.

Roberts, L. G. [1965]. "Machine Perception of Three-Dimensional Solids," in *Optical and Electro-Optical Information Processing*, Tippet, J. T. (ed.), MIT Press, Cambridge, Mass.

Robertson, A. R. [1977]. "The CIE 1976 Color Difference Formulae," *Color Res. Appl.*, vol. 2, pp. 7–11.

Robinson, G. S. [1976]. "Detection and Coding of Edges Using Directional Masks," University of Southern California, Image Processing Institute, Report no. 660.

Robinson, J. A. [1965]. "A Machine-Oriented Logic Based on the Resolution Principle," *J. ACM*, vol. 12, no. 1, pp. 23–41.

Robinson, J. [2006]. "Adaptive Prediction Trees for Image Compression," *IEEE Trans. Image Proc.*, vol. 15, no. 8, pp. 2131–2145.

Rock, I. [1984]. *Perception*, W. H. Freeman, New York.

Roese, J. A., Pratt, W. K., and Robinson, G. S. [1977]. "Interframe Cosine Transform Image Coding," *IEEE Trans. Comm.*, vol. COM-25, pp. 1329–1339.

Rosenblatt, F. [1959]. "Two Theorems of Statistical Separability in the Perceptron," In *Mechanisation of Thought Processes: Proc. of Symposium No. 10*, held at the National Physical Laboratory, November 1958, H. M. Stationery Office, London, vol. 1, pp. 421–456.

Rosenblatt, F. [1962]. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan, Washington, D. C.

Rosenfeld, A. (ed.) [1984]. *Multiresolution Image Processing and Analysis*, Springer-Verlag, New York.

Rosenfeld, A. [1999]. "Image Analysis and Computer Vision: 1998," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 36–95.

Rosenfeld, A. [2000]. "Image Analysis and Computer Vision: 1999," *Computer Vision and Image Understanding*, vol. 78, no. 2, pp. 222–302.

Rosenfeld, A. and Kak, A. C. [1982]. *Digital Picture Processing*, vols. 1 and 2, 2nd ed., Academic Press, New York.

Rosin, P. L. [1997]. "Techniques for Assessing Polygonal Approximations of Curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 6, pp. 659–666.

Rudnick, P. [1966]. "Note on the Calculation of Fourier Series," *Math. Comput.*, vol. 20, pp. 429–430.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. [1986]. "Learning Internal Representations by Error Propagation," In *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 1: Foundations*, Rumelhart, D. E., et al. (eds.), MIT Press, Cambridge, Mass., pp. 318–362.

Rumelhart, D. E. and McClelland, J. L. (eds.) [1986]. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 1: Foundations*, MIT Press, Cambridge, Mass.

Runge, C. [1903]. *Zeit. für Math., and Physik*, vol. 48, p. 433.

Runge, C. [1905]. *Zeit. für Math., and Physik*, vol. 53, p. 117.

Runge, C. and König, H. [1924]. "Die Grundlehren der Mathematischen Wissenschaften," *Vorlesungen iiber Numerisches Rechnen*, vol. 11, Julius Springer, Berlin.

Russ, J. C. [1999]. *The Image Processing Handbook*, 3rd ed., CRC Press, Boca Raton, Fla.

Russo F. and Ramponi, G. [1994]. "Edge Extraction by FIRE Operators," *Fuzz-IEEE '94*, vol. 1, pp. 249–243, IEEE Press, New York.

Sahni, S. and Jenq, J.-F. [1992]. "Serial and Parallel Algorithms for the Medial Axis Transform," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 12, pp. 1218–1224.

Sahoo, S. S. P. K., Wong, A. K. C., and Chen, Y. C. [1988]. "Survey of Thresholding Techniques," *Computer Vision, Graphics and Image Processing*, vol. 41, pp. 233–260.

Saito, N. and Cunningham, M. A. [1990]. "Generalized E-Filter and its Application to

Edge Detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 8, pp. 814–817.
Sakrison, D. J. and Algazi, V. R. [1971]. "Comparison of Line-by-Line and Two-Dimensional Encoding of Random Images," *IEEE Trans. Info. Theory*, vol. IT-17, no. 4, pp. 386–398.
Salari, E. and Siy, P. [1984]. "The Ridge-Seeking Method for Obtaining the Skeleton of Digital Images," *IEEE Trans. Syst. Man Cyb.*, vol. SMC-14, no. 3, pp. 524–528.
Salinas, R. A., Abidi, M. A., and Gonzalez, R. C. [1996]. "Data Fusion: Color Edge Detection and Surface Reconstruction Through Regularization," *IEEE Trans. Industrial Electronics*, vol. 43, no. 3, pp. 355–363, 1996.
Sato, Y. [1992]. "Piecewise Linear Approximation of Plane Curves by Perimeter Optimization," *Pattern Recog.*, vol. 25, no. 12, pp. 1535–1543.
Sauvola, J. and Pietikainen, M. [2000]. "Adaptive Document Image Binarization," *Pattern Recog.*, vol. 33, no. 2, pp. 225–236.
Schalkoff, R. J. [1989]. *Digital Image Processing and Computer Vision*, John Wiley & Sons, New York.
Schonfeld, D. and Goutsias, J. [1991]. "Optimal Morphological Pattern Restoration from Noisy Binary Images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 1, pp. 14–29.
Schowengerdt, R. A. [1983]. *Techniques for Image Processing and Classification in Remote Sensing*, Academic Press, New York.
Schreiber, W. F. [1956]. "The Measurement of Third Order Probability Distributions of Television Signals," *IRE Trans. Info. Theory*, vol. IT-2, pp. 94–105.
Schreiber, W. F. [1967]. "Picture Coding," *Proc. IEEE* (Special issue on redundancy reduction), vol. 55, pp. 320–330.
Schreiber, W. F. and Knapp, C. F. [1958]. "TV Bandwidth Reduction by Digital Coding," *Proc. IRE National Convention*, pt. 4, pp. 88–99.
Schwartz, J. W. and Barker, R. C. [1966]. "Bit-Plane Encoding: A Technique for Source Encoding," *IEEE Trans. Aerosp. Elec. Systems*, vol. AES-2, no. 4, pp. 385–392.
Selesnick, I., Baraniuk, R., and Kingsbury, N. [2005]. "The Dual-Tree Complex Wavelet Transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151.
Serra, J. [1982]. *Image Analysis and Mathematical Morphology*, Academic Press, New York.
Serra, J. (ed.) [1988]. *Image Analysis and Mathematical Morphology*, vol. 2, Academic Press, New York.
Sezan, M. I., Rabbani, M., and Jones, P. W. [1989]. "Progressive Transmission of Images Using a Prediction/Residual Encoding Approach," *Opt. Eng.*, vol. 28, no. 5, pp. 556–564.
Shack, R. V. [1964]. "The Influence of Image Motion and Shutter Operation on the Photographic Transfer Function," *Appl. Opt.*, vol. 3, pp. 1171–1181.
Shafarenko, L., Petrou, M., and Kittler, J. [1998]. "Histogram-Based Segmentation in a Perceptually Uniform Color Space," *IEEE Trans. Image Processing*, vol. 7, no. 9, pp. 1354–1358.
Shaked, D. and Bruckstein, A. M. [1998]. "Pruning Medial Axes," *Computer Vision and Image Understanding*, vol. 69, no. 2, pp. 156–169.
Shannon, C. E. [1948]. "A Mathematical Theory of Communication," *The Bell Sys. Tech. J.*, vol. XXVII, no. 3, pp. 379–423.
Shapiro, L. G. and Stockman, G. C. [2001]. *Computer Vision*, Prentice Hall, Upper Saddle River, N.J.
Shapiro, V. A. [1996]. "On the Hough Transform of Multi-Level Pictures," *Pattern Recog.*, vol. 29, no. 4, pp. 589–602.
Shariat, H. and Price, K. E. [1990]. "Motion Estimation with More Than Two Frames," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 5, pp. 417–434.
Shepp, L. A. and Logan, B. F. [1974]. "The Fourier Reconstruction of a Head Section," *IEEE Trans. Nucl. Sci.*, vol. NS-21, pp. 21–43.
Sheppard, J. J., Jr., Stratton, R. H., and Gazley, C., Jr. [1969]. "Pseudocolor as a Means of Image Enhancement," *Am. J. Optom. Arch. Am. Acad. Optom.*, vol. 46, pp. 735–754.
Shi, F. Y. and Wong, W.-T. [1994]. "Fully Parallel Thinning with Tolerance to Boundary Noise," *Pattern Recog.*, vol. 27, no. 12, pp. 1677–1695.
Shih, F. Y. C. and Mitchell, O. R. [1989]. "Threshold Decomposition of Gray-Scale Mor-

phology into Binary Morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 1, pp. 31–42.

Shirley, P. [2002]. *Fundamentals of Computer Graphics*, A. K. Peters, Natick, MA.

Sid-Ahmed, M. A. [1995]. *Image Processing: Theory, Algorithms, and Architectures*, McGraw-Hill, New York.

Sikora, T. [1997]. "MPEG Digital Video-Coding Standards," *IEEE Signal Processing*, vol. 14, no. 5, pp. 82–99.

Simon, J. C. [1986]. *Patterns and Operators: The Foundations of Data Representations*, McGraw-Hill, New York.

Sklansky, J., Chazin, R. L., and Hansen, B. J. [1972]. "Minimum-Perimeter Polygons of Digitized Silhouettes," *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 260–268.

Sloboda, F., Zatko, B., and Stoer, J. [1998]. "On Approximation of Planar One-Dimensional Continua," in *Advances in Digital and Computational Geometry*, R. Klette, A. Rosenfeld, and F. Sloboda (eds.), Springer, Singapore, pp. 113–160.

Smirnov, A. [1999]. *Processing of Multidimensional Signals*, Springer-Verlag, New York.

Smith, A. R. [1978]. "Color Gamut Transform Pairs," *Proc. SIGGRAPH* '78, published as *Computer Graphics*, vol. 12, no. 3, pp. 12–19.

Smith, J. O., III [2003]. *Mathematics of the Discrete Fourier Transform*, W3K Publishing, CCRMA, Stanford, CA. (Also available online at http://ccrma.stanford.edu/~jos/mdft).

Smith, M. J. T. and Barnwell, T. P. III [1984]. "A Procedure for Building Exact Reconstruction Filter Banks for Subband Coders," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, San Diego, Calif.

Smith, M. J. T. and Barnwell, T. P. III [1986]. "Exact Reconstruction Techniques for Tree-Structured Subband Coders," *IEEE Trans. On Acoust., Speech, and Signal Proc.*, vol. 34, no. 3, pp. 434–441.

Snyder, W. E. and Qi, Hairong [2004]. *Machine Vision*, Cambridge University Press, New York.

Sobel, I. E. [1970]. "Camera Models and Machine Perception," Ph.D. dissertation, Stanford University, Palo Alto, Calif.

Sonka, M., Hlavac, V., and Boyle, R. [1999]. *Image Processing, Analysis, and Machine Vision*, 2nd ed., PWS Publishing, New York.

Snyder, W. E. and Qi, Hairong [2004]. *Machine Vision*, Cambridge University Press, New York.

Soille, P. [2003]. *Morphological Image Analysis: Principles and Applications*, 2nd ed., Springer-Verlag, New York.

Solari, S. [1997]. *Digital Video and Audio Compression*, McGraw-Hill, New York.

Stark, H. (ed.) [1987]. *Image Recovery: Theory and Application*, Academic Press, New York.

Stark, J. A. [2000]. "Adaptive Image Contrast Enhancement Using Generalizations of Histogram Equalization," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 889–896.

Stockham, T. G., Jr. [1972]. "Image Processing in the Context of a Visual Model," *Proc. IEEE*, vol. 60, no. 7, pp. 828–842.

Storer, J. A. and Reif, J. H., eds. [1991]. *Proceedings of DDC '91*, IEEE Computer Society Press, Los Alamitos, Calif.

Strang, G. and Nguyen, T. [1996]. *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, Mass.

Stumpff, K. [1939]. *Tafeln und Aufgaben zur Harmonischen Analyse und Periodogrammrechnung*, Julius Springer, Berlin.

Sussner, P. and Ritter, G. X. [1997]. "Decomposition of Gray-Scale Morphological Templates Using the Rank Method," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 6, pp. 649–658.

Swets, D. L. and Weng, J. [1996]. "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 8, pp. 1831–1836.

Symes, P. D. [2001]. *Video Compression Demystified*, McGraw-Hill, New York.

Sze, T. W. and Yang, Y. H. [1981]. "A Simple Contour Matching Algorithm," *IEEE Trans.*

Pattern Anal. Mach. Intell., vol. PAMI-3, no. 6, pp. 676–678.

Tanaka, E. [1995]. "Theoretical Aspects of Syntactic Pattern Recognition," *Pattern Recog.*, vol. 28, no. 7 pp. 1053–1061.

Tanimoto, S. L. [1979]. "Image Transmission with Gross Information First," *Comput. Graphics Image Proc.*, vol. 9, pp. 72–76.

Tasto, M. and Wintz, P. A. [1971]. "Image Coding by Adaptive Block Quantization," *IEEE Trans. Comm. Tech.*, vol. COM-19, pp. 957–972.

Tasto, M. and Wintz, P. A. [1972]. "A Bound on the Rate-Distortion Function and Application to Images," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 150–159.

Teh, C. H. and Chin, R. T. [1989]. "On the Detection of Dominant Points on Digital Curves," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 8, pp. 859–872.

Theoridis, S. and Konstantinos, K. [2006]. *Pattern Recognition*, 3rd ed., Academic Press, New York.

Thévenaz, P. and Unser, M [2000]. "Optimization of Mutual Information for Multiresolution Image Registration," *IEEE Trans. Image Processing*, vol. 9, no. 12, pp. 2083–2099.

Thomas, L. H. [1963]. "Using a Computer to Solve Problems in Physics," *Application of Digital Computers*, Ginn, Boston.

Thomason, M. G. and Gonzalez, R. C. [1975]. "Syntactic Recognition of Imperfectly Specified Patterns," *IEEE Trans. Comput.*, vol. C-24, no. 1, pp. 93–96.

Thompson, W. B. (ed.) [1989]. Special issue on visual motion, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 5, pp. 449–541.

Thompson, W. B. and Barnard, S. T. [1981]. "Lower-Level Estimation and Interpretation of Visual Motion," *Computer*, vol. 14, no. 8, pp. 20–28.

Thorell, L. G. and Smith, W. J. [1990]. *Using Computer Color Effectively*, Prentice Hall, Upper Saddle River, N.J.

Tian, J. and Wells, R. O., Jr. [1995]. *Vanishing Moments and Wavelet Approximation*, Technical Report CML TR-9501, Computational Mathematics Lab., Rice University, Houston, Texas.

Tizhoosh, H. R. [2000]. "Fuzzy Image Enhancement: An Overview," in *Fuzzy Techniques in Image Processing*, E. Kerre and M. Nachtegael, eds., Springer-Verlag, New York.

Tomita, F., Shirai, Y., and Tsuji, S. [1982]. "Description of Texture by a Structural Analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 183–191.

Topiwala, P. N. (ed.) [1998]. *Wavelet Image and Video Compression*, Kluwer Academic Publishers, Boston, Mass.

Toro, J. and Funt, B. [2007]. "A Multilinear Constraint on Dichromatic Planes for Illumination Estimation," *IEEE Trans. Image Proc.*, vol. 16, no. 1, pp. 92–97.

Tou, J. T. and Gonzalez, R. C. [1974]. *Pattern Recognition Principles*, Addison-Wesley, Reading, Mass.

Tourlakis, G. J. [2003]. *Lectures in Logic and Set Theory*, Cambridge University Press, Cambridge, UK.

Toussaint, G. T. [1982]. "Computational Geometric Problems in Pattern Recognition," In *Pattern Recognition Theory and Applications*, Kittler, J., Fu, K. S., and Pau, L. F. (eds.), Reidel, New York, pp. 73–91.

Tsai, J.-C., Hsieh, C.-H, and Hsu, T.-C. [2000]. "A New Dynamic Finite-State Vector Quantization Algorithm for Image Compression," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1825–1836.

Tsujii, O., Freedman, M. T., and Mun, K. S. [1998]. "Anatomic Region-Based Dynamic Range Compression for Chest Radiographs Using Warping Transformation of Correlated Distribution," *IEEE Trans. Medical Imaging*, vol. 17, no. 3, pp. 407–418.

Udpikar, V. R. and Raina, J. P. [1987]. "BTC Image Coding Using Vector Quantization," *IEEE Trans. Comm.*, vol. COM-35, no. 3, pp. 352–356.

Ueda, N. [2000]. "Optimal Linear Combination of Neural Networks for Improving Classification Performance," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 2, pp. 207–215.

Ullman, S. [1981]. "Analysis of Visual Motion by Biological and Computer Systems,"

IEEE Computer, vol. 14, no. 8, pp. 57–69.
Umbaugh, S. E. [2005]. *Computer Imaging: Digital Image Analysis and Processing*, CRC Press, Boca Raton, FL.
Umeyama, S. [1988]. "An Eigendecomposition Approach to Weighted Graph Matching Problems," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 5, pp. 695–703.
Unser, M. [1995]. "Texture Classification and Segmentation Using Wavelet Frames," *IEEE Trans. on Image Processing*, vol. 4, no. 11, pp. 1549–1560.
Unser, M., Aldroubi, A., and Eden, M. [1993]. "A Family of Polynomial Spline Wavelet Transforms," *Signal Proc.*, vol. 30, no. 2, pp. 141–162.
Unser, M., Aldroubi, A., and Eden, M. [1993]. "B-Spline Signal Processing, Parts I and II," *IEEE Trans. Signal Proc.*, vol. 41, no. 2, pp. 821–848.
Unser, M., Aldroubi, A., and Eden, M. [1995]. "Enlargement or Reduction of Digital Images with Minimum Loss of Information," *IEEE Trans. Image Processing*, vol. 4, no. 5, pp. 247–257.
Vaidyanathan, P. P. and Hoang, P.-Q. [1988]. "Lattice Structures for Optimal Design and Robust Implementaion of Two-Channel Perfect Reconstruction Filter Banks," *IEEE Trans. Acoust., Speech, and Signal Proc.*, vol. 36, no. 1, pp. 81–94.
Vailaya, A., Jain, A. and Zhang, H. J. [1998]. "On Image Classification: City Images vs. Landscapes," *Pattern Recog.*, vol. 31, no. 12, pp. 1921–1935.
Vetterli, M. [1986]. "Filter Banks Allowing Perfect Reconstruction," *Signal Proc.*, vol. 10, no. 3, pp. 219–244.
Vetterli, M. and Kovacevic, J. [1995]. *Wavelets and Suband Coding*, Prentice Hall, Englewood Cliffs, N.J.
Vincent, L. [1993]. "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms," *IEEE Trans. Image Proc.*, vol. 2. no. 2, pp. 176–201.
Voss, K. and Suesse, H. [1997]. "Invariant Fitting of Planar Objects by Primitives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 1, pp. 80–84.
Vuylsteke, P. and Kittler, J. [1990]. "Edge-Labeling Using Dictionary-Based Relaxation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, no. 2, pp. 165–181.
Walsh, J. W. T. [1958]. *Photometry*, Dover, New York.
Wang, D., Zhang, L., Vincent, A., and Speranza, F. [2006]. "Curved Wavelet Transform for Image Coding," *IEEE Trans. Image Proc.*, vol. 15, no. 8, pp. 2413–2421.
Wang, G., Zhang, J., and Pan, G.-W. [1995]. "Solution of Inverse Problems in Image Processing by Wavelet Expansion," *IEEE Trans. Image Processing*, vol. 4, no. 5, pp. 579–593.
Wang, Y.-P., Lee, S. L., and Toraichi, K. [1999]. "Multiscale Curvature-Based Shape Representation Using β-Spline Wavelets," *IEEE Trans. Image Processing*, vol. 8, no. 11, pp. 1586–1592.
Wang, Z., Rao, K. R., and Ben-Arie, J. [1996]. "Optimal Ramp Edge Detection Using Expansion Matching," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 11, pp. 1092–1097.
Watt, A. [1993]. *3D Computer Graphics*, 2nd ed., Addison-Wesley, Reading, Mass.
Wechsler [1980]. "Texture Analysis – A Survey," *Signal Proc.*, vol. 2, pp. 271–280.
Wei, D., Tian, J., Wells, R. O., Jr., and Burrus, C. S. [1998]. "A New Class of Biorthogonal Wavelet Systems for Image Transform Coding," *IEEE Trans. Image Processing*, vol. 7, no. 7, pp. 1000–1013.
Weinberger, M. J., Seroussi, G., and Sapiro, G. [2000]. "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS," *IEEE Trans. Image Processing*, vol. 9, no. 8, pp. 1309–1324.
Westenberg, M. A. and Roerdink, J. B. T. M. [2000]. "Frequency Domain Volume Rendering by the Wavelet X-Ray Transform," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1249–1261.
Weszka, J. S. [1978]. "A Survey of Threshold Selection Techniques," *Comput. Graphics Image Proc.*, vol. 7, pp. 259–265.
White, J. M. and Rohrer, G. D. [1983]. "Image Thresholding for Optical Character Recognition and Other Applications Requiring Character Image Extraction," *IBM*

J. Res. Devel., vol. 27, no. 4, pp. 400–411.
Widrow, B. [1962]. "Generalization and Information Storage in Networks of 'Adaline' Neurons," In *Self-Organizing Systems 1962*, Yovitz, M. C., et al. (eds.), Spartan, Washington, D. C., pp. 435–461.
Widrow, B. and Hoff, M. E. [1960]. "Adaptive Switching Circuits," *1960 IRE WESCON Convention Record, Part 4*, pp. 96–104.
Widrow, B. and Stearns, S. D. [1985]. *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, N.J.
Wiener, N. [1942]. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, the MIT Press, Cambridge, Mass.
Wilburn, J. B. [1998]. "Developments in Generalized Ranked-Order Filters," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 15, no. 5, pp. 1084–1099.
Windyga, P. S. [2001]. "Fast Impulsive Noise Removal," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 173–179.
Wintz, P. A. [1972]. "Transform Picture Coding," *Proc. IEEE*, vol. 60, no. 7, pp. 809–820.
Witten, I. H., Neal, R. M., and Cleary, J. G. [1987]. "Arithmetic Coding for Data Compression," *Comm. ACM*, vol. 30, no. 6, pp. 520–540.
Wolberg, G. [1990]. *Digital Image Warping*, IEEE Computer Society Press, Los Alamitos, CA.
Wolff, R. S. and Yaeger, L. [1993]. *Visualization of Natural Phenomena*, Springer-Verlag, New York.
Won, C. S. and Gray, R. M. [2004]. *Stochastic Image Processing*, Kluwer Academic/Plenum Publishers, New York.
Woods, J. W. and O'Neil, S. D. [1986]. "Subband Coding of Images," *IEEE Trans. Acous. Speech Signal Proc.*, vol. ASSP-35, no. 5, pp. 1278–1288.
Woods, R. E. and Gonzalez, R. C. [1981]. "Real-Time Digital Image Enhancement," *Proc. IEEE*, vol. 69, no. 5, pp. 643–654.
Xu, Y., Weaver, J. B., Healy, D. M., Jr., and Lu, J. [1994]. "Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique," *IEEE Trans. Image Processing*, vol. 3, no. 6, pp. 747–758.
Xu, R., Pattanaik, S., and Hughes, C. [2005]. "High-Dynamic-Range Still-Image Encoding in JPEG 2000," *IEEE Computer Graphics and Applications*, vol 25, no. 6, pp. 57–64.
Yachida, M. [1983]. "Determining Velocity Maps by Spatio-Temporal Neighborhoods from Image Sequences," *Comput. Vis. Graph. Image Proc.*, vol. 21, no. 2, pp. 262–279.
Yamazaki, Y., Wakahara, Y., and Teramura, H. [1976]. "Digital Facsimile Equipment 'Quick-FAX' Using a New Redundancy Reduction Technique," *NTC '76*, pp. 6.2-1–6.2-5.
Yan, Y. and Cosman, P. [2003]. "Fast and Memory Efficient Text Image Compression with JBIG2," *IEER Trans. Image Proc.*, vol. 12, no. 8, pp. 944–956.
Yang, X. and Ramchandran, K. [2000]. "Scalable Wavelet Video Coding Using Aliasing-Reduced Hierarchical Motion Compensation," *IEEE Trans. Image Processing*, vol. 9, no. 5, pp. 778–791.
Yates, F. [1937]. "The Design and Analysis of Factorial Experiments," Commonwealth Agricultural Bureaux, Farnam Royal, Burks, England.
Yin, P. Y., Yin, L. H., and Chen, L. H. [1997]. "A Fast Iterative Scheme for Multilevel Thresholding Methods," *Signal Processing*, vol. 60, pp. 305–313.
Yitzhaky, Y., Lantzman, A., and Kopeika, N. S. [1998]. "Direct Method for Restoration of Motion Blurred Images," *J. Opt. Soc. Amer.-A. Optics, Image Science, and Vision*, vol. 15, no. 6, pp. 1512–1519.
You, J. and Bhattacharya, P. [2000]. "A Wavelet-Based Coarse-to-Fine Image Matching Scheme in a Parallel Virtual Machine Environment," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1547–1559.
Yu, D. and Yan, H. [2001]. "Reconstruction of Broken Handwritten Digits Based on Structural Morphology," *Pattern Recog.*, vol. 34, no. 2, pp. 235–254.
Yu, S. S. and Tsai, W. H. [1990]. "A New Thinning Algorithm for Gray-Scale Images,"

Pattern Recog., vol. 23, no. 10, pp. 1067–1076.
Yuan, M. and Li, J. [1987]. "A Production System for LSI Chip Anatomizing," *Pattern Recog. Letters*, vol. 5, no. 3, pp. 227–232.
Zadeh, L. A. [1965]. "Fuzzy Sets," *Inform. and Control*, vol. 8, pp. 338–353.
Zadeh, L. A. [1973]. "Outline of New Approach to the Analysis of Complex Systems and Decision Processes," *IEEE Trans. Systems, Man, Cyb.*, vol. SMC-3, no. 1, pp. 28–44.
Zadeh, L. A. [1976]. "A Fuzzy-Algorithmic Approach to the Definition of Complex or Imprecise Concepts," *Int. J. Man-Machine Studies*, vol. 8, pp. 249–291.
Zahara, E., Shu-Kai, S., and Du-Ming, T. [2005]. "Optimal Multi-Thresholding Using a Hybrid Optimization Approach," *Pattern Recognition Letters*, vol. 26, no. 8, pp. 1082–1095.
Zahn, C. T. and Roskies, R. Z. [1972]. "Fourier Descriptors for Plane Closed Curves," *IEEE Trans. Comput.*, vol. C-21, no. 3, pp. 269–281.
Zhang, T. Y. and Suen, C. Y. [1984]. "A Fast Parallel Algorithm for Thinning Digital Patterns," *Comm. ACM*, vol. 27, no. 3, pp. 236–239.
Zhang, Y. and Rockett, P. I. [2006]. "The Bayesian Operating Point of the Canny Edge Detector," *IEEE Trans. Image Proc.*, vol. 15, no. 11, pp. 3409–3416.
Zhu, H., Chan F. H. Y., and Lam, F. K. [1999]. "Image Contrast Enhancement by Constrained Local Histogram Equalization," *Computer Vision and Image Understanding*, vol. 73, no. 2, pp. 281–290.
Zhu, P. and Chirlian, P. M. [1995]. "On Critical Point Detection of Digital Shapes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 8, pp. 737–748.
Zimmer, Y., Tepper, R., and Akselrod, S. [1997]. "An Improved Method to Compute the Convex Hull of a Shape in a Binary Image," *Pattern Recog.*, vol. 30, no. 3, pp. 397–402.
Ziou, D. [2001]. "The Influence of Edge Direction on the Estimation of Edge Contrast and Orientation," *Pattern Recog.*, vol. 34, no. 4, pp. 855–863.
Ziv, J. and Lempel, A. [1977]. "A Universal Algorithm for Sequential Data Compression," *IEEE Trans. Info. Theory*, vol. IT-23, no. 3, pp. 337–343.
Ziv, J. and Lempel, A. [1978]. "Compression of Individual Sequences Via Variable-Rate Coding," *IEEE Trans. Info. Theory*, vol. IT-24, no. 5, pp. 530–536.
Zucker, S. W. [1976]. "Region Growing: Childhood and Adolescence," *Comput. Graphics Image Proc.*, vol. 5, pp. 382–399.
Zugaj, D. and Lattuati, V. [1998]. "A New Approach of Color Images Segmentation Based on Fusing Region and Edge Segmentation Outputs," *Pattern Recog.*, vol. 31, no. 2, pp. 105–113.

索　引

A

Accumulative difference images (ADIs)　累积差分图像　801–802
Achromatic (monochromatic) light　无色(单色)光　67, 418
Acoustic imaging　声波成像　42–44
Acquisition　获取　见 *Image acquisition*
Adaptive context dependent probability　自适应上下文相关概率　572–573
Adaptive filters　自适应滤波器　见 *Spatial filters*
Additive cost functions　加性代价函数　537
Additivity　加性　95, 366
Adjacency of pixels　邻接像素　90–91
Affine transformations　仿射变换　109–111　见 *Geometric transformations*
Aggregation of fuzzy sets　模糊集合的聚类　204, 207
Aliasing　混淆(混叠)　239–241, 250–257
　filtering and　滤波和～　240, 251
　image interpolation and resampling and　图像内插、重取样和～　252–255
　moiré patterns and　莫尔(波纹)模式和～　255–257
　spatial　空间～　251
　temporal　时间～　251
Alpha-trimmed mean filter　修正的阿尔法均值滤波器　349–352
Analysis filter banks　分析滤波器组　492, 525–526
Analysis trees, wavelet packets　分析树，小波包　532–536
Anti-aliasing　反混淆(混叠)　240, 251
Approximation coefficients　近似系数　494, 508, 511
Approximation pyramids　近似金字塔　486–488
Arithmetic coding　算术编码　570–573
Arithmetic logic unit (ALU)　算术逻辑单元(ALU)　51
Arithmetic mean filter　算术均值滤波器　344
Arithmetic operations　算术操作　96–102
Array operations　阵列操作　94–95
Autocorrelation　自相关　375
Autocorrelation matrix　自相关矩阵　621
AVS compression　AVS 压缩　560, 563

B

Back propagation, neural network training by　反向传播，神经网络训练的　914–921
Background　背景　92, 105
Backprojections　反投影　385–387
　fan-filtered　扇形滤波～　403–409
　filtered　滤波～　397–400, 403–409
　halo-effect blurring from　～引起的晕环效应模糊　385–387
　parallel-beam filtered　平行射线束滤波～　397–403
Band-limited functions　带限函数　236–239, 249–250
Bandpass filters　带通滤波器　316, 358, 412
Bandreject filters　带阻滤波器　316, 357, 412
Bartlane cable system　沃特兰电缆系统　25–26
Basis functions　基函数　499, 589, 592–593
　DCT　离散余弦变换～　591
　Haar　哈尔～　496
　series expansion using　使用～的级数展开　499
　Walsh-Hadamard　沃尔什-哈达玛～　590
Basis images　基图像　见 *Basis functions*
Bayes　贝叶斯
　classification　～分类　896–904
　classifier　～分类器　895
　decision function　～决策函数　896–898
　decision rule　～决策规则　764
　formula　～公式　766
Bidirectional frames (B-frames)　双向帧(B 帧)　612
Binary images　二值图像　90, 650
　border following　～的边界追踪　818
　boundary of　～的边界　92
　compression of　～的压缩　576, 584
　logical operations on　对～的逻辑操作　105
　morphological operations on　对～的形态学操作　650–686
　segmentation and　分割和～　465, 718, 748, 796
Binary trees　二叉树　532
Biorthogonality　双正交性　492
Bit-plane coding　比特(位)平面编码　584–588
Bit-plane slicing　比特平面分层(分割)　139
Bit rate　比特率　559
Bits　比特　52, 80–81, 82
Blind　盲
　deconvolution　～去卷积　368
　spot　～点　59
Block matching　块匹配　612–613
Block transform coding　块变换编码　588–606
　bit allocation for　～的比特分配　596–601
　JPEG compression and　JPEG 压缩和～　601–606
　selection of transform for　为～选取变换　589–595
　subimage size and　子图像大小和～　595–596
　threshold implementation　～阈值实现　599–601
　zonal implementation　～分区实现　598–599
Blurring　模糊　见 *Filtering*
BMP compression　BMP 压缩　560, 563, 576
Border　边界　92　见 *Boundary*
　clearing　～清除　685–686
　following　～追踪　818–820
　inner　内～　92
　outer　外～　92
Bottom-hat transformation　底帽变换　694–696
Boundary　边界　见 *Border, Regional descriptors*
　definition　～的定义　92
　chain codes　～链码　820
　curvature of　～的曲率　837–838
　decomposition of　～的分解　832–834
　description　～的描述　837–844
　detection of segmentation　～的分割检测　747–760
　diameter　～直径　837
　eccentricity of　～偏心率　837
　edge linking and　边缘连接和～　747–760
　extraction　～提取　211, 664–665
　following　～追踪　818–820
　Fourier descriptors for　～的傅里叶描绘子　840–843
　length　～长度　837
　Moore boundary tracking algorithm　摩尔边界追踪算法　818–819
　pixels　～像素　92–93
　polygonal approximation　～的多边形近似　823–830
　representation　～表示　817–837
　segments　～线段　832–834
　signatures　～的标记图　830–832
　shape numbers of　～的形状数　838–839
　statistical moments of　～的统计矩　843–844

Brightness　亮度　61–65, 67, 418, 420
　adaptation of human eye　人眼的～适应性　61–65
　chromatic light and　彩色光和～　67, 418
　color image processing and　彩色图像处理和～　418, 420
　subjective　主观～　61–62
Butterworth filters　布特沃斯滤波器
　bandpass　带通～　316, 358
　bandreject　带阻～　316, 357
　highpass（BHPF）　高通～　306–307
　lowpass（BLPF）　低通～　295–298, 373
　notch　陷波～　317, 359
　sharpening using　使用～锐化　306–307
　smoothing using　使用～平滑　295–298

C

Canny edge detector　坎尼边缘检测器　741–747
Cartesian product　笛卡儿积　79, 203, 687
CAT　计算机轴向断层　见 *Computed tomography*
Cataracts　白内障　59
CCD arrays　CCD 阵列　60, 72, 81, 335, 414, 473
CCITT　国际电话电报咨询委员会　560
CCITT compression　CCITT 压缩　578–581
Chain codes　链码　820–823
Chessboard distance　棋盘距离　93
Chromatic（color）light　彩色光　67, 418
Chromaticity diagram　色度图　421–422
City-block distance　城市街区距离　93
Classifiers　分类器
　Bayes　贝叶斯～　896–904
　minimum distance　最小距离～　888–891
　neural network　神经网络～　904–924
　optimum statistical　最佳统计～　894–904
　probability and　概率和～　895–896
　structural　结构～　925–928
Closing　闭操作　657–661, 690–692, 699
　gray-scale morphology and　灰度级形态学和～　690–692, 699
　morphological operation of　～的形态学操作　657–661
　reconstruction, by　～重建　699
CMY color model　CMY 彩色模型　424, 428–429
CMYK color model　CMYK 彩色模型　424, 429
Code　编码（名词）　见 *Compression*
　arithmetic　算术～　570–573
　block　块～　565
　CCITT makeup　CCITT 补偿～　934
　CCITT terminating　CCITT 终结～　933
　Elias gamma　Elias gamma～　569
　Golomb　Golomb～　566–570
　Gray　格雷～　585
　Huffman　霍夫曼～　564–566
　JPEG default AC　JPEG 默认 AC～　935–936
　JPEG default DC　JPEG 默认 DC～　935
　instantaneous　瞬时～　565
　length　长度～　549
Code　编码（动词）
　MH（Modified Huffman）coding　MH（改进的霍夫曼）～　577
　MMR（modified modified READ）　MME（改进的改进的 READ）～　578
　MR（modified READ）　MR（改进的 READ）～　578
　natural binary　自然二进制～　550
　READ（relative element address designate）　相对元素寻址指派～　578
　Rice　莱斯～　567
　symbols　符号～　549
　unary　一元～　566
　uniquely decodable　唯一可解码的～　565
　variable-length　变长～　551
　words　码字　549
Codec　编解码　558
Coding　编码　488–495, 549, 550–552, 562, 564–636　见 *Compression*
　methods for image compression　图像压缩的～方法　562, 564–636
　redundancy　～冗余　549, 550–552
　subband　子带～　488–495
　symbol-based（or token-based）　基于符号（或基于标记）的～　581–584
Cohen-Daubechies-Feauveau biorthogonal wavelets
　Cohen-Daubechies-Feauveau 双正交小波　540–541
Color　彩色（颜色）
　fundamentals　～基础　417
　gamut　～色域　422
　models　～模型　423–436
　safe browser　稳定浏览器～　426
　safe RGB　稳定 RGB～　426
　safe（Web）　稳定（Web）～　426
Color image processing　彩色图像处理　416–482
　chromaticity diagram　色度图　421–422
　color corrections　彩色校正　455
　color "gradient"　彩色梯度　471
　CMY model　CMY 模型　424, 428–429
　CMYK model　CMYK 模型　424, 429
　color slicing　彩色分层（分割）　453
　compression and　压缩和～　476–477
　edge detection　边缘检测　469
　full-color processing　全彩色处理　416, 446–448
　histogram processing　直方图处理　460
　HSI model　HIS 模型　424, 429–436
　intensity slicing　灰度分层　437
　intensity to color　灰度至彩色　440
　models for　～的模型　423–436
　noise in　～中的噪声　473–476
　pseudocolor　伪彩色　416, 436–446
　RGB model　RGB 模型　423–424, 424–428
　segmentation　分割　467–472
　sharpening　（尖）锐化　464–465
　smoothing in　～中的平滑　461–464
　transformations in　～中的变换　448–461
　trichromatic coefficients　三色系数　421
Color transformations　彩色变换　448–461
　color circle for　～的彩色环　452
　color management systems（CMS）for　～的彩色管理系统　455–459
　complements　～的补色　452–453
　corrections to color and tone　～中颜色和色调的校正　455–459
　formulation for　～的公式　448–451
　histogram processing for　～的直方图处理　460–461
　profiles for　～的剖面线　455–456
　slicing　～分层（分割）　453–455
　tonal range for　～的色调范围　456–458
Commission Internationale del'Eclairage（CIE）　国际照明委员会　419, 421–422, 456
Compact support　紧支撑　503
Complex numbers　复数　224–225
Compression　压缩　49, 476–477, 547–648
　arithmetic coding　算术编码～　570–573
　bit-plane coding　比特平面编码～　584–588
　block transform coding　块变换编码～　588–606
　BMP　位图（扩展名）　576
　CCITT　577–581
　coding redundancy　编码冗余～　549, 550–551
　color images　彩色图像～　476–477
　containers for　～的容器（存储器）　560–562, 563
　fidelity criteria　～的保真度准则　556–558
　formats for　～的格式　560–562, 563
　fundamentals of　～的基础　548–562
　Golomb coding　Golomb 编码　566–570
　Huffman coding　霍夫曼编码　564–566

irrelevant information and 不相关的信息和~ 549, 552–553
JBIG-2 JBIG-2~ 583–588
JPEG JPEG~ 601–606
JPEG-2000 JPEG-2000~ 629–635
Lempel-Ziv-Welch (LZW) coding LZW 编码 573–575
mapping and 映射和~ 552, 559–560
measuring information for ~的度量信息 553–556
methods of ~的方法 562, 564–636
models for ~的模型 558–560
MPEG-4 AVC (H.264) MPEG-4 AVC (H.264)~ 616–618
predictive coding 预测编码~ 606–625
quantization and 量化和~ 553, 559–560, 618–620, 624–625
ratio ~率 548–549
run-length coding 行程编码~ 575–581
spatial redundancy 空间冗余~ 549, 551–552
standards for ~的标准 560–562, 563
symbol-based coding 基于符号的编码~ 581–584
temporal redundancy 时间冗余~ 549, 551–552
wavelet coding 小波编码~ 626–636
Components of image processing system 图像处理系统的组成 50–52
Computed tomography (CT) 计算机断层 28, 33, 71, 334, 384–409
Computerized axial tomography (CAT) 计算机轴向断层 见 *Computed tomography*
Connected component 连通分量
definition ~的定义 91
description ~的描述 845–849
extraction of ~的提取 667–669, 707
segmentation ~的分割 786, 794
Connected pixels 连通像素 91
Connected set 连通集 91
Constrained least squares filtering 约束最小平方滤波 379–383
Containers for image compression 图像压缩的存储器 560–562, 563
Continuous wavelet transform (CWT) 连续小波变换 513–515
scale and translation in ~中的尺度和平移 513
admissibility criterion 容许准则 513
Contour 轮廓 见 *Border, Boundary*
Contraharmonic mean filter Contraharmonic 均值滤波器 345–347
Contrast 对比度 24, 80, 100, 119, 142, 208, 869 见 *Enhancement*
local 局部~ 780
medium 中等~ 99, 139
measure of ~度量 850, 854–856
simultaneous 瞬时~ 63
stretching ~拉伸 128, 137, 138
Control points 控制点 112
Convex hull 凸壳
definition ~的定义 669
extraction ~的提取 669–671
for description ~的描述 832–834
Convex deficiency 凸缺 669
Convolution 卷积
by digital filtering 数字滤波~ 489
circular 循环~ 245, 271
filter ~滤波器 172
integral 积分~ 367
kernel ~核 172
mask ~模板 172
spatial continuous 空间连续~ 231–32, 433
spatial discrete 空间离散~ 168–172
theorem ~定理 232, 271, 276, 285, 367, 401, 811, 892
Co-occurrence matrix 共生矩阵 852–858
Correlation 相关
circular 循环~ 276
coefficient ~系数 642, 892
descriptor ~描绘子 853, 856
matching by ~匹配 891–894
spatial 空间~ 168–172
theorem ~定理 277
Cross-modulation 交叉调制 492
CT 见 *Computed tomography*
Cutoff frequency 截止频率 292

D

Dam construction for watersheds 分水岭的水坝构建 794–796
Data compression 数据压缩 548 见 *Compression*
Dead zones 死区 629
Decimation 抽取 253
Decision function 决策函数 888
Decision surfaces, complexity of 决策面, ~的复杂度 921–924
Decoding 解码 558, 560
Huffman coding and 霍夫曼编码和~ 565
image decompression and 图像解压缩和~ 558, 560
inverse mapper for ~的反映射器 560
symbol decoder for ~的符号解码器 560
Decomposition 分解 537–540, 628–629
boundary segments from ~导致的边界分割 832–834
level selection for wavelet coding 小波编码的~级别选取 628–629
trees in wavelet packets 小波包中的~树 537–540
wavelets and 小波和~ 537–540, 628–629
Defense Meteorological Satellite Program (DMSP) 国防气象卫星计划 37
Defuzzification 去模糊 204–205, 207
Degradation 退化 见 *Restoration*
estimating ~估计 368–372
linear, position-invariant 线性位(置)不变~ 365–368
model of ~的模型 334–335
Delta modulation (DM) 德尔塔(增量)调制 619–620
Denoising 去噪 334, 530
Derivative 导数 见 Gradient, Laplacian
first order 一阶~ 180–182, 715
second order 二阶~ 180–182, 715
Description 描述 837–877
area 面积~ 837
basic rectangle 基本矩形~ 837
boundary 边界~ 837
circularity ratio 圆度率~ 844
compactness 致密性~ 844
diameter 直径~ 837
eccentricity 偏心~ 837
Euler number 欧拉数~ 845
Fourier descriptors 傅里叶描绘子~ 840
moment invariants 矩不变~ 861
perimeter 周长~ 844
principal components 主成分~ 864
regional 区域~ 见 *Regional descriptors*
relational 关系~ 874
shape numbers 形状数~ 838
statistical moments 统计矩~ 843
texture 纹理~ 849–861
topological 拓扑~ 845
Denoising 去噪 530
Detail coefficients (horizontal, vertical, and diagonal) 细节系数(水平、垂直和对角线) 494, 508, 511
Differential pulse code modulation (DPCM) 差分脉冲编码调制 621–624
Digital 数字
filter ~滤波器 见 *Filters*
image, definition of ~图像, ~图像的定义 23
Digital image processing 数字图像处理 见 *Image*
defined ~的定义 23–25
fields of ~的领域 29–47
fundamentals of ~基础知识 57–125
high-level processes of ~的高级处理 24
history of ~的历史 25–29

origins of ~的起源 25–29
sensors for ~的传感器 50, 68–73
steps in ~的步骤 47–50
Digital signal filtering 数字信号滤波 488–491
Digital signal processing (DSP) 数字信号处理 488–491
Digital Video Disk (DVD) 数字视频光盘 547–548
Digitizer 数字化器 50, 70
Dilation 膨胀 见 *Morphological image processing*
Dilation equation 膨胀公式 504
Discrete cosine transform (DCT) 离散余弦变换 591 见 *JPEG compression*
Discrete Fourier transform (DFT) 离散傅里叶变换
average value 平均值 268, 275
circular convolution 循环卷积 见 *Convolution*
circular correlation 循环相关 见 *Correlation*
derivation of ~的导数 224–235
Fast Fourier Transform (FFT) 快速傅里叶变换 321–325
implementation ~的实现 320–325
padding ~的填充 273–275
pair, 1-D ~一维对 258
periodicity of ~的周期性 259–261
phase angle ~的相角 267, 275
polar representation ~极坐标表示 275
properties ~属性 258–275
separability ~的可分离性 276
spectrum ~谱 229, 248, 267, 275
symmetry properties ~的对称性 264
two-dimensional 二维~ 257–258
zero padding ~的零填充 273–274
wraparound error ~的缠绕错误 272
Discrete wavelet transform (DWT) 离散小波变换 510–512, 524 见 *Wavelets*
Discriminant (decision) analysis 判别(决策)分析 884–885, 888
Distance measures 距离度量(测度) 93–94, 114–115, 467, 784–785, 831, 837, 888–891, 899, 925
Dots (pixels) 点(像素)
per inch (DPI) ~每英寸 81, 256, 581
per unit distance ~每单位距离 81
Downsampling 下取(采)样 486–487
DPI 点每英寸 81, 256, 581
DV compression DV 压缩 560, 562
Dynamic range 动态范围 79–80

E

Edge 边缘 见 *Edge detection*
color 彩色~ 469–472
definition ~的定义 92
direction ~的方向 728
enhancement ~增强 179–190, 302–311, 693
gradient ~梯度 187, 471, 623, 693, 728
linking ~连接 747–760
magnitude ~幅度 187–188, 728
map ~映射 733
models ~模型 722–728
noise sensitivity ~噪声敏感性 726–727
normal ~法线 729
operators ~算子 730
ramp 斜坡~ 181, 715, 724
roof 屋顶~ 715, 724
step 台阶~ 181, 715, 724
types ~类型 180–182, 716
unit normal 单位法线~ 729
wavelet transform and 小波变换和~ 526–527, 529–530
zero crossing ~零交叉 181, 725, 739
Edge detection 边缘检测 469–472, 722–747 见 *Edge*
boundary detection 边界检测 747
Canny edge detector 坎尼边缘检测器 741–747
derivatives 导数 180–184, 715–716
edge linking 边缘连接 747–760
false negative 伪负值 744
false positive 伪正值 744
gradient 梯度 187, 471, 623, 693, 728–736 见 *Gradient*
gradient and thresholding 梯度和阈值处理 735
hysteresis thresholding 滞后阈值处理 744
Laplacian of Gaussian (LoG) 高斯拉普拉斯 737
Marr-Hildreth edge detector Marr-Hildreth 边缘检测器 736–741
models for ~的模型 722–728
nonmaxima suppression 非最大抑制~ 743
Prewitt edge detector Prewitt 边缘检测器 730–732, 809
ramp edges 斜坡边缘 715–717, 722
Roberts detector 罗伯特检测器 189, 730
roof edges 屋顶边缘 715, 723–724
Sobel edge detector Sobel 边缘检测器 188–190, 730–732, 810
spaghetti effect 意大利通心粉效应 739
spatial filters and 空间滤波器和~ 717
step edges 台阶边缘 715–717, 722
wavelet-based 基于小波的~ 529–530
Electromagnetic (EM) spectrum 电磁波谱 24, 29–42, 65–68
gamma radiation 伽马辐射 30–31, 67–68
imaging in ~图像处理 29–42
importance of ~的重要性 24
infrared regions 红外区域 34–40, 68
light and 光和~ 65–68
microwave band 微波频段 40–42, 67, 68
radio band 无线电频段 42, 67, 68
source of image from 来自~的图像源 29–30
units of ~的单位 66, 67
visible band 可见光频段 34–40, 66–67
X-rays X 射线 31–33, 67–68
Electron beam computed tomography 电子束计算机断层 389
Electron microscopy 电子显微镜 29, 42, 68, 137, 164, 278
Elias gamma codes Elias gamma 编码 569
Encoding 编码 558, 559, 575–577 见 *Compression*
image compression and 图像压缩和~ 558, 559
mapper for ~映射器 559
quantizer for ~量化器 559
run-length (RLE) 行程~ 575–577
symbol coder for ~的符号编码器 559
Empty set 空集 102
Enhancement 增强
adaptive 自适应~ 150, 352, 354
contrast enhancement 对比度增强 135, 149, 150, 208, 311, 332
contrast stretching 对比度拉伸 128, 137, 138
combined methods 组合的方法 191–195
defined 定义的~ 47, 129, 223
frequency domain 频率域~ 279–320
fuzzy techniques for ~的模糊技术 208–213
homomorphic filtering 同态滤波 311
image averaging 图像平均 97
image subtraction 图像相减 99
histogram processing for ~的直方图处理 142–166
intensity transformations 灰度变换 129–141
local 局部~ 161, 164, 352, 354
median filter 中值滤波器 178, 217, 348, 354, 411
order statistic filters 统计排序滤波器 178, 347
sharpening (尖)锐化 179, 302
smoothing 平滑 97, 174, 291
spatial filters 空间滤波器 166–190
Entropy 熵 554–555
Erlang (gamma) noise 爱尔兰(伽马)噪声 337–338
Erosion 腐蚀 见 *Morphological image processing*
Estimating the degradation function 估计退化函数 368–372
Euclidean distance 欧氏距离 114 见 *Distance*
measures ~度量(测度)
norm ~范数 114
Expansions 展开 499–508, 508–510

basis functions of ~的基函数 499
biorthogonal 双正交~ 500
coefficients of ~的系数 499
multiresolution analysis (MRA) 多分辨率分析 499, 503–504
orthonormal 归一化正交~ 500
overcomplete 过完备~ 500
scaling functions ~尺度函数 499, 501–505
series 级数~ 499–501, 508–510
wavelet functions for ~的小波函数 505–508
wavelet series 小波级数~ 508–510
Exponential Golomb codes 指数 Golomb 编码 569
Exponential noise 指数噪声 338

F

False color 假彩色 见 *Pseudocolor*
False contouring 假轮廓 85, 122, 141, 645
Fan-beam filtered backprojections 扇形射线束滤波反投影 403–409
Fast Fourier transform (FFT) 快速傅里叶变换
Fast wavelet transform (FWT) 快速小波变换 515–523, 524–527, 532–541
analysis filter banks 分析滤波器组 517–518, 525–526
image compression using 使用~的图像压缩 626–635
inverse 反~ 520–522
multi-resolution processing using 使用~的多分辨率处理 515–523, 524–527
synthesis filter banks 综合滤波器组 521–522, 525–526
time-frequency tiles 时间-频率片 522–523
two-dimensional 二维~ 523–527
wavelet packets for ~的小波包 532–541
FAX 传真 577
Feature selection 特征选择 见 *Description*
Fidelity criteria 保真度准则 556–558
Fiducial marks 基准标记 117
Filters 滤波器
deconvolution 去卷积~ 368
frequency domain 频率域~
kernels ~的核 见 *Spatial filters*
finite impulse response (FIR) 有限冲激响应~ 286, 490
Hamming window ~的汉明窗 399
Hann window ~的汉宁窗 399
reconstruction ~重建 239
spatial 空间~ 见 *Spatial filters, Spatial filtering* ~滤波
transfer function ~的传输函数 279
zero-phase-shift 零相移~ 284
Filter banks 滤波器组 491–493
Filters, digital 数字滤波器 488–495
biorthogonal 双正交~ 492, 540–541
coefficients ~系数 490
Cohen-Daubechies-Feauveau biorthogonal coefficients Cohen-Daubechies-Feauveau 双正交系数 540
convolution and 卷积和~ 489
Daubechies 8-tap orthonormal coefficients 8 抽头 Daubechies 正交归一化系数 494
filter banks 滤波器组 491–493
filter taps 滤波器抽头 490
finite impulse response 有限冲激响应 490
FIR 有限冲激响应 490
Haar coefficients 哈尔系数 519
impulse response 冲激响应 490
JPEG-2000 irreversible JPEG-2000 不可逆 31–29, 631
modulation in ~中的调制 491
order of ~的阶 490
order reversal in ~的反阶 491
orthonormal 归一化正交~ 493–494, 519, 529
perfect reconstruction 完美重建 492
prototypes 原型 493
sign reversal in ~的符号反转 490
symlet (4th order orthonormal) coefficients 对称小波(第 4 阶正交归一化)系数 529
Filter banks 滤波器组 491–493
FWT analysis FWT 分析~ 517–520, 533
FWT synthesis FWT 合成~ 521–522
wavelet packet analysis 小波包分析 535
Filtering 滤波
frequency 频率~ 见 *Frequency domain filtering*
spatial 空间~ 见 *Spatial filtering*
Finite impulse response (FIR) filters 有限冲激响应滤波器 286, 490
Fixed increment correction rule 固定增量修正规则 908
Fluorescence microscopy 荧光显微镜 33–34
Foreground 前景 92, 105
Formats for image compression 图像压缩格式 560–562, 563
Forward mapping 前向映射 109
Fourier descriptors 傅里叶描绘子 840–843
Fourier series 傅里叶级数 222–223, 225
Fourier-slice theorem 傅里叶切片定理 396–397
Fourier spectrum 傅里叶谱 131–132, 228–229, 267–271
log transformations and 对数变换和~ 131–132
phase angle and 相角和~ 267–271
plot of frequency of ~的频率图 228–229
Fourier transform 傅里叶变换 227–277
continuous 连续~ 227, 248
convolution 卷积 见 *Convolution*
discrete 离散~ 见 *Discrete Fourier transform*
Fast Fourier transform (FFT) 快速傅里叶变换 见 *Discrete Fourier transform*
history of ~的历史 222–223, 326
pair ~对 117, 227, 232, 244, 248, 258, 892
power spectrum ~的功率谱 267
sampling and 取样和~ 233–242, 249–257
Fractal images 分形图像 46–47
Frame buffers 帧缓存器 52
Freeman chain code 费雷曼链码 820–823
Frequency domain 频率域 221–332, 804–807
additional characteristics ~的附加特征 277–279
aliasing ~混淆(混叠) 见 *Aliasing*
convolution ~卷积 见 *Convolution*
discrete Fourier transform (DFT) ~离散傅里叶变换 见 *Discrete Fourier transform*
fast Fourier transform (FFT) ~快速傅里叶变换 见 *Discrete Fourier transform*
filtering ~滤波 见 *Frequency domain*
Fourier series ~傅里叶级数 222–223, 225
Fourier spectrum ~傅里叶谱 267–271
Fourier transform ~傅里叶变换 见 *Fourier transform*
impulse ~冲激 见 Impulse
motion in segmentation ~分割中的运动 804–807
sampling ~取样 见 Sampling
sifting property ~取样特性 见 impulse
Frequency domain filtering 频率域滤波 277–320. 见 *Spatial filtering*
bandpass filters 带通滤波器 316–320, 357–362
bandreject filters 带阻滤波器 316–320, 357–362
box filter 盒状滤波器 229
Butterworth filters 布特沃斯滤波器 295–298, 306–307, 316–319, 357–360, 373
correspondence with spatial filtering 与空间滤波的对应性 285, 291
fundamentals of ~基础 279–285
Gaussian filters for ~的高斯滤波器 280–281, 287–291, 298–299, 307–308, 316–319, 357–360
highboost filters 高提升滤波器 310
high frequency emphasis 高频强调 310
highpass filters for ~的高通滤波器 280, 303–308

homomorphic filters 同态滤波器 311–315
ideal filters 理想滤波器 238–239, 250, 282–284, 291–295, 299, 303–307, 316, 357–360
Laplacian 拉普拉斯 308–310
lowpass filters 低通滤波器 239, 280, 291–303
notch filters 陷波滤波器 316–320, 357–362
sharpening (尖)锐化 303–315
smoothing 平滑 291–303
steps ~的步骤 285
unsharp masking 非锐化模板 310
Frequency intervals 频率间隔 245–246
Frequency spectrum 频谱 见 *Spectrum*
FWT 518, 533
subband coding 子带编码 491
wavelet packet 小波包 535–536
Front-end subsystem 前端子系统 51
Full-color image processing 全彩色图像处理 416, 446–448
Functionally complete 功能完备的 105
Fuzzy sets 模糊集合 106–107, 195–213
aggregation of ~的聚集 204, 207
color fuzzified by 被~模糊的彩色 200–208
definitions for ~定义 196–200
defuzzification of ~的去模糊 204–205, 207
implication of ~的含义 201–204, 207
intensity transformations and 灰度变换和~ 208–211
membership (characteristic) functions 隶属度(特征)函数 106, 195–200
principles of theory 理论原理 196–200
set operations of ~的集合操作 106–107, 195–196
spatial filtering and 空间滤波和~ 208–213
use of ~的用途 200–208, 208–211, 211–213

G
Gamma 伽马
correction ~校正 133–135
noise ~噪声 见 *Noise*
Gamma-ray imaging 伽马射线成像 30, 43, 69
Gaussian filter 高斯滤波器
frequency 频率 见 *Frequency domain filtering*
spatial 空间 见 *Spatial filtering*
Gaussian noise 高斯噪声 见 *Noise*
Gaussian pattern class 高斯模式类 896–904
Gaussian pyramid 高斯金字塔 486
Geometric mean filter 几何均值滤波器 345, 383–384
Geometric transformations 几何变换 109–114
Affine 仿射 109
control points ~控制点 112
identity 恒等~ 110
rotation 旋转~ 110
scaling 尺度~ 110
shearing 剪切~ 110
tie points ~连接点 112
translation ~平移 110
GIF compression GIF 压缩 560, 563, 573
Global thresholding 全局阈值处理 见 *Thresholding*
Golomb codes and coding Glomb 代码和编码 566–570
Golomb-Rice codes Golomb-Rice 编码 567
Gradient 梯度 187–190, 469–473, 693–694, 728–736
color segmentation 彩色分割 469–473
edge detection 边缘检测 728–736
edge normal (vector) 边缘法线(向量) 729
edges 边缘 190, 469–473
first-order derivatives, as 一阶导数 187–190
gray-scale morphology 灰度级形态学 693–694
morphological 形态学~ 693
operators ~算子 188–190, 469–473, 729–734
Prewitt operators Prewitt 算子 731–732
properties of ~的属性 728–729
Roberts operators Roberts 算子 188–189, 730–730
sharpening (尖)锐化 187–190
Sobel operators Sobel 算子 188–190, 731–732
thresholding, combined with 阈值处理与~ 735–736
Granular noise 颗粒噪声 620
Granulometry 粒度 696–697
Gray level 灰度级 23, 67, 74, 128 见 *Intensity*
Gray level co-occurrence matrix 灰度级共生矩阵 852 见 *Intensity*
Gray scale 灰度级 67, 74
Gray-scale morphology 灰度级形态学 687–701 见 *Morphological image processing*
bottom-hat transformation 底帽变换 694–696
closing 闭操作 690–692, 699
dilation 膨胀 688–690, 698–699
erosion 腐蚀 688–690, 699
gradient 梯度 693–694
granulometry 粒度 696–697
opening 开操作 690–692, 699
reconstruction 重建 698–701
smoothing 平滑 692–693
textural segmentation 纹理分割 697–698
top-hat transformation 顶帽变换 694–696

H
Haar transform 哈尔变换 496–499
Halftone dots 半色调点 256–257
Hamming window 汉明窗 399
Hann window 汉宁窗 399
Harmonic mean filter 同态均值滤波器 345
HDV compression HDV 压缩 560, 563
Heisenberg cells/boxes 海森伯单元/盒 522
Heisenberg uncertainty principle 海森伯测不准原理 522
Hertz (Hz) 赫兹 66
High definition (HD) television 高清晰度电视 548
High-frequency-emphasis filtering 高频强调滤波 310–311
Highboost filtering 高提升滤波 184–187, 310–311
Highpass filters 高通滤波器
frequency ~频率 见 *Frequency domain filtering*
spatial 空间~ 见 *Spatial filtering*
HSI color model HSI 彩色模型 424, 429–436, 465–467
conversion from RGB RGB 至~的转换 432–433
conversion to RGB ~至 RGB 的转换 433–435
manipulation of images 图像加工 435–436
plane concept of ~的平面概念 430–432
segmentation ~分割 465–467
uses of ~的用途 429
Histogram processing 直方图处理 142–166, 460–461
definition ~的定义 142
color transformation using 使用~的彩色变换 460–461
equalization 均衡~ 144–150
global 全局~ 142–160
intensity transformation 灰度变换 144, 148
inverse transformation 反变换 144, 150
local 局部~ 161–166
matching (specification) 匹配(规定化) 150–160
normalized 归一化~ 142
probability density function (PDF) for ~的概率密度函数 145–147
statistics, use of 统计~, ~的用途 161–166
Hit-or-miss transformation 击中或击不中变换 662–663
Hole filling 孔洞填充 665–667, 682, 684–685, 707
Homogeniety 一致性 95, 366, 854
Homomorphic filtering 同态滤波 311–315
Hough transform 霍夫变换 755–760
Hue, color image processing and 色调、彩色图像处理和~ 420–421, 429–436
Huffman coding 霍夫曼编码 564–566
Human eye 人眼 见 *Visual perception*

H.261, H.262, H.263, H.264 560, 562, 616–618

I
Ideal filter 理想滤波器 见 *Frequency domain filtering*
IEC 国际电工委员会 560
Illumination 照明 73–74, 762–763
  correction ~校正 100–101, 694–695, 778, 783
  eye response 人眼响应~ 39, 59, 62
  image model, in ~中的图像模型 73–74, 311–315
  nonuniform 不均匀~ 100–101, 694–695, 763, 778
  segmentation and 分割与~ 762–763
  source ~源 68–72
  standard ~标准 456, 630
  structured light 结构光~ 39
Image 图像
  acquisition ~获取 68–72
  analysis ~分析 24
  blur 模糊~ 369–372
  color processing 彩色~处理 416–482
  compression ~压缩 见 *Compression*
  deconvolution ~去卷积 368
  element ~元素 见 Pixel
  enhancement ~增强 见 *Enhancement*
  filtering ~滤波 见 *Filtering*
  formation model ~的形成模型 72, 311
  illumination ~照明 见 *Illumination*
  intensity ~亮度 见 *Intensity*
  interpolation ~内插 见 *Interpolation*
  morphology ~形态学 见 *Morphological image processing*
  pixel ~像素 见 *Pixel*
  reflectance ~反射 73, 311
  registration ~配准 97, 111, 801, 864
  resampling ~重取样 87, 252, 639, 821
  restoration ~复原 见 *Image restoration*
  rotation ~旋转 见 *Geometric transformations*
  scaling ~尺度变换 见 *Geometric transformations*
  segmentation ~分割 见 *Segmentation*
  sensing 遥感~ 29–47, 68–72
  sensors ~传感器 68–72
  shearing ~剪切 见 *Geometric transformations*
  translation ~平移 见 *Geometric transformations*
  zooming ~缩放 87, 109, 252
Image compression standards 图像压缩标准 560–563
Image file formats and image containers 图像文件格式和图像容器 560–563
Image information 图像信息 553–556
Image pyramids 图像金字塔 485–488
Image transforms 图像变换 见 *Transforms*
Imaging modalities 成像模态 30–47
Implication in fuzzy sets 模糊集合中的含义 201–204, 207
Impulse 冲激
  continuous 连续~ 225–226, 247–248
  discrete 离散~ 169–171, 247–248
  noise ~噪声 178–179, 338–340
  response ~响应 286, 366–367, 369, 490, 494, 631
  sifting property of ~的取样特性 225–227, 247–248, 490
  train ~串 226–227, 230–231, 250
  unit discrete 单位离散~ 169–171, 226, 247–248
Independent frames (I-frames) 独立帧(I帧) 611
Information theory 信息论 554–556
Infrared 红外 29, 34, 43, 66, 99, 418, 440, 444, 712, 845, 849, 868, 901
Intensity 灰度 23, 67, 81–87
  fuzzy techniques ~模糊技术 195, 208–211
  mean ~均值 见 *Moments*
  mapping ~映射 564, 109–111, 128–166, 448
  quantization ~量化 74–76
  scale ~尺度 74
  scaling ~尺度变换 101–102
  statistical descriptors ~统计描绘子 118–119, 161–166
  transformations ~变换 107, 127–166
  thresholding ~阈值处理 760–785
  transformations ~变换 128–166
  variance ~方差 162 见 *Moments*
Intensity transformations 灰度变换 128
  bit-plane slicing 比特平面分层~ 139
  contrast stretching 对比度拉伸~ 128, 137, 138
  gamma 伽马~ 132
  histogram equalization 直方图均衡~ 142–150
  histogram matching 直方图匹配~ 150–160
  histogram specification 直方图规定化 150–160
  intensity-level slicing 灰度级分层 137
  local 局部~ 161–166
  log 对数~ 131
  negative 反转~ 130
  piecewise linear 分段线性~ 137
  power law 幂律~ 132
Interpolation 内插 87–90, 109–113, 242, 252–255, 485, 562, 615
  bicubic 双立方~ 88
  bilinear 双线性~ 88
  nearest neighbor 最近邻~ 87–88
  resampling (shrinking and zooming) images by ~重取样(缩放)图像 87–90
Inverse filtering 逆滤波 373–374
Inverse Fourier transform 傅里叶反变换 见 *Fourier transform, Discrete Fourier transform*
Inverse mapping 反(逆)映射 109
Inverse transforms 反变换 见 *Transforms*
Invisible watermarks 不可见水印 638–642
ISO 国际标准化组织 560
Isopreference curves 等偏爱曲线 86
Isotropic filters 各向同性滤波器 182
ITU-T 国际电信联盟远程通信标准化组织 560

J
Jaggies 锯齿 254
JBIG compression JBIG压缩 560, 561
JBIG2 compression JBIG2压缩 560, 561, 583–584
JPEG compression JPEG压缩 560, 561, 601–606, 629–636
  block transform coding for ~的块变换编码 601–606
JPEG-2000 standard JPEG-2000标准 629–636
  wavelet coding for ~的小波编码 629–636
JPEG-LS compression JPEG-LS压缩 560, 561, 572
  JPEG-2000 compression JPEG-2000压缩 560, 561, 629–635
  components ~成分 630
  derived vsexpounded quantization 导出式量化与描述式量化 633
  irreversible component transform 不可逆成分变换 630
  lifting-based wavelet transforms 基于提升的小波变换 631
  tile components 片成分 631

L
LANDSAT satellite LANDSAT卫星 36, 806, 848
Laplacian 拉普拉斯
  defined ~的定义 182
  color 彩色~ 464
  convolution using 使用~的卷积 811
  combined with gradient ~与梯度的组合 191, 772
  decomposition ~分解 812
  frequency domain 频率域~ 277, 308, 329–330
  isotropic property ~的各向同性 219, 721
  of Gaussian (LoG) 高斯~ 737, 811
  operators ~算子 183
  PDF ~概率密度函数 610
  pyramid 金字塔~ 488
  restoration for ~的复原 380
  scaling ~尺度变换 184
  sharpening with 使用~锐化 184–185, 309

thresholding for　~的阈值处理　718–721, 736, 771–775
zero crossing　~零交叉　181, 725, 739
Large scale integration（LI）大规模集成　27
Least-mean-square（LMS）delta rule　最小均方德尔塔规则　909
Lempel-Ziv-Welch（LZW）coding　Lempel-Ziv-Welch 编码　573–575
Light　光　65–68, 417–423　见 *Electromagnetic（EM）spectrum*
absorption of　~的吸收　418–419
achromatic　无色~　418
chromatic　色~　418
color image processing and　彩色图像处理和~　417–423
microscopy　显微镜　35
monochromatic　单色~　67
vision and　视觉与~　见 *Visual perception*
EM spectrum visible band for　~的 EM 谱可见波段　65–68, 417–418
primary and secondary color of　~的原色和二次色　419–420
Line detection　线检测　719–722
Line pairs　线对
per mm　~每毫米　81
per unit distance　每单位距离　81
Linear　线性
convolution　~卷积　见 *Convolution*
correlation　~相关　见 *Correlation*
FIR filters　FIR 滤波器　286
frequency domain filters　频率域滤波器　272
masks　模板　172
motion　~运动　371, 388
operations　~操作　95–96, 276, 365–368
transforms　~变换　115
spatial filters　~空间滤波器　167, 172
system　~系统　225, 334, 365–368
Linearly separable classes　线性可分类　908–909
Live image　活动图像　99
Lloyd-Max quantizer　劳埃德-马克斯量化器　625
Log transformations　对数变换　131–132
Logical operations　逻辑操作（运算）　105–106
Lossless predictive coding　无损预测编码　606–611
Lossy predictive coding　有损预测编码　618–621
Lowpass filters　低通滤波器
frequency　~频率　见 *Frequency domain*
filtering　~滤波
spatial　空间~　见 *Spatial filtering*
LSB watermarks　LSB 水印　638
Luminance, chromatic light and　亮度，色光和~　67, 418
LZW coding　LZW 编码　见 *Lempel-Ziv-Welch（LZW）coding*

M
Mach bands　马赫带　63, 64
Macroblocks　宏块　611
Magnetic resonance imaging（MRI）　磁共振成像　42, 72, 112, 135, 390
Mahalanobis distance　马氏距离　785　见 *Distance measures*
Mallat's herringbone algorithm　Mallat herringbone 算法　515
Mapper　映射器　559
Mapping　映射　109–110, 154–155, 157–158, 552, 559–560　见 *Intensity*
decoding（decompression）and　解码（分解）和~　560
encoding（compression）and　编码（压缩）和~　559
forward　前向~　109–110
histogram processing and　直方图处理和~　154–155, 157–158
inverse　反~　110, 560
Markers　标记
morphological reconstruction for　~的同态重建　678–686, 698–699
thresholding　~阈值处理　772
watersheds for　分水岭~　798–800
Markov sources　马尔可夫源　556
Marr-Hildreth edge detector　Marr-Hildreth 边缘检测器　736–741
Masks　模板　见 Spatial filters
definition　~的定义　128
masking function　模板函数　593
threshold　~阈值　599
unsharp masking and　非锐化模板和~　184–187
Mask mode radiography　遮光方式造影　99
Matching　匹配　888–894, 925–928
block　块~　612–613
correlation, by　~相关　891–894
minimum distance classifier　最小距离分类器
method　~方法　888–891
shape numbers　~形状数　925–926
strings　~串　926–928
Matrix operations　矩阵操作　78, 94–95, 114–115
array operations versus　阵列操作与~　94–95
notation for pixels　~像素的表示　78
vector operations and　向量运算和~　114–115
Max filters　最大滤波器　174, 348
Mean absolute distortion（MAD）　均值绝对失真　612
Mean filters　均值滤波器　见 *Spatial filters*
Mean of intensity　亮度均值　见 *Moments*
Mean pyramid　均值金字塔　486
Mean square error（MSE）　均方误差
filtering in　~的滤波　374–379
measure　~度量　376
Medial axis transformation（MAT）　中轴变换　834–835
Median filters　中值滤波器　178–179, 348, 411
adaptive　自适应~　354–357
updating　更新~　218
Membership（characteristic）functions　隶属度（特征）函数　106, 195–200
Mexican hat　墨西哥帽
operator　~算子　737
wavelet　~小波　514
Micron　微　66
Microdensitometer　微光密度计计　70
Microwave　微波　29, 40, 66, 440
Midpoint filter　中点滤波器　349
Min filter　最小滤波器　179, 349
Minimum distance classifier　最小距离分类器　888–891
Minimum-perimeter polygon（MPP）　最小周长多边形　823–829
Minkowsky　闵可夫斯基
addition　~相加　705
subtraction　~相减　704
M-JPEG　560, 563
Modified Huffman（MH）coding　改进的霍夫曼编码　577
Modified READ（MR）coding　改进的 READ 编码　578
Modified Modified READ（MMR）coding　改进的改进的 READ 编码　578
Modulation　调制　491
Modulation function　调制函数　363
Moiré patterns　莫尔（波纹）模式　255–257, 318
Moments　矩
statistical　统计~　118–119, 843, 850, 881, 885
invariant　不变~　861–864
Monochromatic（achromatic）light　单色（无色）光　67, 418
Moore boundary tracking algorithm　摩尔边界追踪算法　818–819
Morphological image processing　形态学图像处理　649–710
alternating sequential filtering　交替顺序滤波~　692
binary images, summary　二值图像，小结　684–686
black top-hat　黑色顶帽~　694
border clearing　边界清除　见 *Morphological reconstruction*
bottom-hat transformation　底帽变换　694
boundary extraction　边界提取　664–665
closing　闭操作　657–391, 690–692
connected components　连通分量　667–669
convex hull　凸壳　669–671
dilation　膨胀　655–657, 678–681, 688–690
erosion　腐蚀　652–655, 657, 678–681, 688–690

filtering 滤波 649, 655, 660, 692, 709
gradient 梯度 693
granulometry 粒度 696
gray-scale 灰度 687–702
hit-or-miss transformation 击中或击不中变换 662–663
hole filling 孔洞填充 665–667, 684–685
opening 开操作 657–661, 681, 684, 690–692
operations summary of ~的操作小结 684–686
preliminaries ~预备知识 650–652
pruning 裁剪 676–678
reconstruction ~的重建 见 *Morphological reconstruction*
reflection of sets in ~中的集合的反射 650
set operations for ~的集合操作 102–106, 650–652
shading correction 阴影校正 695
skeletons 骨架 673–676 见 *Skeletons*
smoothing 平滑 692
structuring element 结构元 651
textural segmentation 纹理分割 697
thickening 粗化 672–673
thinning 细化 671–672
top-hat transformations 顶帽变换 694, 699
translation of sets in ~的集合平移 651
white top-hat 白色顶帽 694
Morphological reconstruction 形态学重建 678–686, 698–701
border clearing and 边界清除和~ 685–686
dilation by 被~膨胀 680–681, 698–699
erosion by 被~腐蚀 680–681, 699
geodesic dilation and erosion 测地膨胀和腐蚀 678–681, 698–699
gray-scale images and 灰度图像和~ 698–701
hole filling and 孔洞填充和~ 684–685
opening by 被~开操作 681, 684, 699
top-hat by 被~顶帽 699
Motion compensation, predictive coding and 运动补偿，预测编码和~ 611–618
Motion estimation 运动估计 612–616
Motion in segmentation 分割中的运动 800–807
accumulative difference image (ADI) 累积差分图像 801–802
frequency domain techniques for 频率域技术 804–807
reference images, establishment of 参考图像，~的建立 803–804
spatial techniques for ~的空间技术 800–804
Moving averages for thresholding 阈值处理的移动平均 781–783
MPEG-1, MPEG-2, MPEG-4 (AVC) 560, 562, 616–618
MQ-coder MQ 编码器 572
Multilayer feedforward neural networks 多层前馈神经网络 841–924
Multiresolution analysis (MRA) 多分辨率分析 499, 503–504
requirements for ~的需求 503–504
Multiresolution processing 多分辨率处理 483–546
expansions 展开 499–508
Haar transform 哈尔变换 496–499
image pyramids 图像金字塔 485–488
MRA equation MRA 方程 504
multiresolution analysis (MRA) 多分辨率分析 499, 503–504
scaling functions 尺度函数 499, 501–505, 523–524
series expansions 级数展开 499–501, 508–510
subband coding 子带编码 488–495
theory of ~的理论 483–484
wavelets and 小波和~ 483–546
Multispectral imaging 多光谱成像 36–37, 114, 444, 848, 868–871, 901–903

N

Nanometer 纳米 66
Negative images 反转图像(负像) 104, 107, 130–131
Neighborhood 邻域
definition ~的定义 90
operations ~操作 107–109, 127–128, 167–191
Neighbor of a pixel 像素的邻域 90
nearest 最近~ 88, 109–111, 242, 252 见 *Interpolation*
types ~类型 90–91
Neural networks 神经网络 904–924
algorithms for ~算法 908–911
back propagation, training by 反向传播，~训练 914–921
background of ~的背景 904–905
decision surfaces, complexity of 决策面，~复杂性 921–924
multilayer feedforward 多层前馈 841–924
perceptrons for ~的感知机 905–907, 908–911
training (learning) process for ~的训练(学习)处理 904–924
training patterns ~训练模式 904
N-largest coding 最大 *N* 编码 599
Noise 噪声 75, 80, 161
bipolar 双极~ 338
color images in ~中的彩色图像 473
data-drop-out 数据清除 338
Erlang 爱尔兰~ 337
exponential 指数~ 338
gamma 伽马~ 337
Gaussian 高斯~ 98, 336
impulse 脉冲~ 178, 338
models ~模型 335
parameter estimation ~参数估计 341
periodic 周期~ 319, 340–341, 357
power spectrum 功率谱~ 375
probability density functions (PDF) 概率密度函数 336–341
Rayleigh 瑞利~ 336
reduction 降低~ 见 *Filtering*
salt-and-pepper 椒盐~ 178, 338
spatial and frequency properties of ~的空间和频率属性 335–336
spike ~尖峰 338
uniform 均匀~ 338
unipolar 单极~ 338
white 白~ 335, 376, 530, 742, 806
Noiseless coding theorem 无噪声编码理论 555
Nonlinear 非线性
filtering ~滤波 167, 174, 178, 187, 347, 352, 892
operation ~操作 95–96, 124
Nonseparable classes 不可分类 909–911
Notch filters 陷波滤波器 见 *Frequency domain filtering*
Null set 空集 102
Nyquist rate 奈奎斯特率 见 *Sampling*

O

Object recognition 对象识别 见 *Patterns, Recognition*
Opening 开操作 见 *Morphological image processing*
Optical illusions 错觉 64–65
Order-statistic filters 统计排序滤波器 见 *Spatial filters*
Ordered pairs product 顺序对积 见 *Cartesian*
Orthonormality 归一化正交 493
Otsu's method Otsu 方法 见 *Threshold, Thresholding*

P

Parallel-beam filtered backprojections 平行射线束滤波反投影 397–403
Parallel distributed processing (PDP) models 并行分布处理模型 904
Patterns 模式 883–931
back propagation and 反向传播与~ 914–921
class structure of ~类结构 883–887
classifiers ~分类器 888–891, 894–904
decision surfaces and 决策面和~ 921–924
discriminant (decision) analysis for ~的判别(决策)分析 884–885, 887
Gaussian class 高斯类~ 896–904
linearly separable classes 线性可分类~ 908–909
matching ~匹配 888–894, 925–928
multiclass recognition 多类识别~ 911–924
neural networks and 神经网络与~ 904–924
nonseparable classes 不可分类~ 909–911

object recognition and 对象识别与～ 883–924
perceptrons and 感知与～ 905–907, 908–841
recognition and 识别与～ 883–931
training (learning) ～训练(学习) 904–924
vector generation for ～的向量生成 884–886
PDF 概率密度函数 560, 563, 585
Pel 像素 见 *Pixel*
Percentile 百分比 179, 348–349, 773
Perceptrons 感知机 905–907, 908–841
Perfect reconstruction filters 完美重建滤波器 492–493
Periodic impulses 周期冲激 见 *Impulse train*
Phase angle 相角 见 *Fourier transform, Discrete Fourier transform*
Photoconverter 光电转换器 69
Photodiode 光电二极管 70
Photons 光子 29, 67
Photopic vision 白昼(亮)视觉 59
Piecewise-linear transformation functions 分段线性变换函数 137–141
Pixel 像素
adjacency of ～邻接 90
array operations ～的阵列操作 94
connected 连通～ 91
definition ～的定义 24, 78
distance between ～间的距离 93
interpolation ～内插 见 *Interpolation*
neighborhood operations filtering 邻域操作滤波 107–109 见 *Spatial*
neighbors of ～的邻域 90
path ～路径 91
per unit distance 每单位距离～ 81
relationships between ～间的关系 90
single operation ～的单个操作 107
transformation ～的变换 见 *Intensity transformations*
PNG compression PNG 压缩 560, 563, 573
Point detection 点检测 见 *Segmentation*
Point processing 点处理 128–129
Point spread function 点扩散函数 367
Polygonal approximation 多边形近似 823–829, 829–830
merging techniques ～聚合技术 829–830
minimum-perimeter polygons (MPP) 最小周长多边形 823–829
splitting techniques ～分裂技术 830
Positron emission tomography (PET) 正电子发射断层术 31, 72, 112, 315, 390, 410
Power-law (gamma) transformations 幂律(伽马)变换 132–137
Power spectrum 功率谱 267, 375
Prediction errors 预测误差 606
Prediction residuals 预测残差 610
motion compensated 运动补偿～ 611–617
pyramid ～金字塔 486, 488
Predictive coding 预测编码 606–625
delta modulation (DM) 德尔塔(增量)调制 619–620
differential pulse code modulation (DPCM) 差分脉冲编码调制 621–624
lossless 无损～ 606–611
lossy 有损～ 618–621
motion compensation and 运动补偿和～ 611–618
optimal predictors for ～的最佳预测器 621–624
optimal quantization in ～中的最佳量化 624–625
prediction error for ～的预测误差 606–607, 621–624
Predictive frames (P-frames) 预测帧(P 帧) 612
Previous pixel predictor 前像素预测器 608
Prewitt gradient operators Prewitt 梯度算子 见 *Spatial filters*
Probability density function (PDF) 概率密度函数 145–147, 336–341, 895–904
Erlang 爱尔兰～ 337
exponential 指数～ 338
gamma 伽马～ 337
Gaussian 高斯～ 98, 336, 897
impulse 冲激～ 178, 338
parameter estimation ～参数估计 341
Rayleigh 瑞利～ 336
salt-and-pepper 椒盐～ 178, 338
uniform 均匀～ 338
Probability mass function (PMF) 概率质量函数 567
Probability models 概率模型 572–573
Projections, image reconstruction from 投影，图像重建形式 384–409
Pruning 裁剪 见 *Morphological image processing*
Pseudocolor image processing 伪彩色图像处理 416, 436–446
intensity slicing for ～的灰度分层 437–440
intensity-to-color transformations 灰度至彩色变换 440–443
monochrome images and 单色图像和～ 444–446
transformations of ～的变换 436–446

Q

Q-coder Q 编码器 572
Quantization 量化 74–90, 553, 559–560, 618–620, 624–625, 629 见 *Sampling*
dead zone ～死区 629
intensity resolution and 灰度分辨率和～ 81–87
interpolation and 内插和～ 87–90
Lloyd-Max quantizer 劳埃德-马克斯量化器 625
mapping and 映射和～ 553, 559–560
optimal 最佳～ 624–625
predictive coding and 预测编码和～ 618–620, 624–625
wavelet coding design of ～的小波编码设计 629
Quicktime 一种压缩格式 560, 563

R

Radiance, chromatic light and 辐射、色光和～ 67, 418
Radio band 无线电波段 29, 42, 66, 301
Ram-Lak filter Ram-Lak 滤波器 398
Random fields 随机场 120
Radon transform 随机变换 388, 390–395
Ramp edges Ramp 边缘 见 *Edges*
Rayleigh noise 瑞利噪声 见 *Noise*
Recognition 识别 49–50, 883–931
Bayes classifier 贝叶斯分类器～ 894–904
classifiers for ～的分类器 888–891, 894–904
correlation 相关～ 891–885
correlation coefficient 相关系数～ 892–894
decision-theoretic methods for ～的决策理论方法 888–924
discriminant analysis 判别分析～ 884
feature selection 特征选取～ 885
learning ～学习 883
matching and 匹配和～ 888–894, 925–928
minimum-distance 最小距离～ 888
neural networks for ～的神经网络 904–924
optimum classifiers 最佳分类器～ 894–896
patterns 模式～ 883–924
shape number matching 形状数匹配～ 925–926
string matching 字符串匹配～ 926–928
structural methods for ～的结构方法 925–928
Reconstruction 重建 239, 241–242, 384–409, 678–686, 699–702
backprojection 反投影～ 385–387, 397–403, 403–409
computed tomography (CT) 计算机断层～ 387–390
fan-beam filtered backprojections 扇形射线束滤波后的反投影～ 403–409
filters 滤波器～ 239
Fourier-slice theorem for ～的傅里叶切片定理 396–397
function, recovery of a 函数，～的恢复 241–242
gray-scale morphological 灰度形态学～ 699–702
image restoration by ～图像复原 384–409
laminogram 层图～ 395
morphological 形态学～ 678–686, 699–702
parallel-beam filtered 平行射线束滤波～
backprojections 反投影～ 397–403
projections, from 投影，来自～ 384–409

Radon transform for ~的雷登变换 390–395
Ram-Lak filter Ram-Lak 滤波器~ 398
Shepp-Logan phantom Shepp-Logan 幻影~ 394
sinogram 正弦图~ 393
Redundancy 冗余 548–552
coding 编码~ 549, 550–551
relative data 相对数据~ 548–549
spatial 空间~ 549, 551–552
temporal 时间~ 549, 551–552
Reference images 参考图像 111–113, 800–804, 806
Refinement equation 细化方程 504
Reflectance 反射 67, 73–74, 311–315, 762–763
Region 区域
definition ~的定义 91
growing ~生长 见 *Region-based segmentation*
of interest (ROI) 感兴趣~ 100, 633, 665, 677, 790
quadregions 1/4~ 789
splitting ~分裂 见 *Region-based segmentation*
descriptors ~描绘子 见 *Description*
Region-based segmentation 基于区域的分割 785–791
merging regions 聚合区域 788–791
region growing 区域生长 785–788
splitting regions 分裂区域 788–791
Regional descriptors 区域描绘子 844–864
area 区域 844
circularity ratio for ~的圆度率 844–845
compactness and 致密性和~ 844–845
contrast ~对比 854–856
correlation ~相关 854–856
entropy ~的熵 854–856
Euler number 欧拉数 847
gray-level co-occurrence matrix 灰度级共生矩阵 852
homogeneity 同质性 854–856
maximum probability 最大概率 854–856
moment invariants for ~的矩不变 861–864
perimeter 周长 844
principal components 主成分 864
relational descriptors 关系描绘子 874
texture content of ~的纹理内容 849–861
topological 拓扑 845–849
uniformity 均匀性, 一致性 854–856
Registration, image 配准, 图像 97, 111, 801, 864
Relative Element Address Designate (READ) coding 相对元素寻址设计编码 578
Remote sensing 遥感 36–37, 548, 893, 901
Representation 表示 49, 817–882
boundary (border) following 边界(边界)追踪~ 818–820
boundary segments for ~的边界分段 832–834
chain codes for ~的链编码 820–823
description and 描述和~ 817–882
polygonal approximation 多边形近似~ 823–829, 829–830
signatures for ~的标记图 830–832
skeletons ~的骨架 834–837
Resampling 重取样 见 *Image resampling*
Reseau marks 网状标记 112
Restoration 复原 48, 333–415
blind deconvolution 盲去卷积~ 368
constrained least squares filtering 约束最小二乘方滤波~ 379–383
deconvolution 去卷积~ 368
degradation functions, estimation 退化函数, 估计 368–373
degradation of an image 一幅图像的退化 333, 334–335, 365–368, 368–373
frequency domain filtering for ~的频率域滤波
noise reduction 降噪 357–365
geometric mean filter 几何均值滤波器 383–384
inverse filtering 逆滤波 373–374
least square error filter 最小平方误差滤波器 375
linear, positive-invariant 线性~, 位置不变~
degradations 退化 365–368
minimum mean square error filtering 最小均方误差滤波 374–379
noise models for ~的噪声模型 335–343
noise reduction and 降噪和~ 344–357, 357–365
parametric Wiener filter 参数维纳滤波器 384
reconstruction ~的重建 见 *Reconstruction*
spatial filtering for noise 噪声的空间滤波
reduction 降低~ 344–357
spectrum equalization filter 谱均衡滤波器 384
Wiener filtering 维纳滤波器 374–379
RGB color models RGB 彩色模型 423–424, 424–428, 432–435, 467–469
conversion from HSI format 从 HIS 格式至 RGB 格式的转换 433–435
conversion to HSI format 从 RGB 格式至 HIS 格式的转换 432–433
cube concept of ~的立方体概念 424–428
safe colors 稳定色 426–428
segmentation and 分割和~ 467–469
Rice codes 莱斯编码 567
Roberts cross-gradient operators 罗伯特交叉梯度算子 188–189, 730–730
Robust invisible watermarks 鲁棒不可见水印 639
Roof edges 屋顶边缘 715, 723–724
Root-mean-square (rms) error 均方根误差 376, 556–558
Rubber-sheet transformations 橡皮膜变换 109–114
Run-length coding (RLE) 行程编码 552, 575–581
Run-length pairs 行程对 552, 575

S

Safe colors 稳定色 426–428
Salt-and-pepper noise 椒盐噪声 见 *Noise*
Sampling 取样 74–90, 233–242, 245–246, 249–257 见 *Quantization*
aliasing ~混淆 见 *Aliasing*
basic concepts of ~的基本概念 74–76
decimation ~抽取 253
Fourier transform and 傅里叶变换和~ 233–242, 249–257
intensity resolution ~灰度分辨率 81–87
interpolation and 内插和~ 87–90, 252–255
intervals ~间隔 245–246
jaggies ~锯齿 254
moiré patterns from ~导致的莫尔模式 255–257
Nyquist rate 奈奎斯特率 237–238
one-variable functions 单变量函数 233–242
reconstruction (recovery) 重建(恢复) 241–242, 252–255
representing digital images by 由~表示数字图像 77–81
sensor arrangement and 传感器排列和~ 76
spatial coordinates (x, y) and 空间坐标(x, y)和~ 74–90
spatial resolution 空间分辨率 81–87
super-sampling 超取样 253
theorem ~定理 235–239, 249–250
two-variable (2-D) functions 二变量(二维)函数 249–257
Saturation 饱和度 80, 320–421
Scaling 尺度变换
geometric 几何~ 见 *Geometric transformations*
intensity 灰度~ 101–102
Scaling functions 尺度函数 499, 501–505, 523–524
coefficients of ~的系数 504
Haar 哈尔~ 502
separable 2D 可分二维~ 523
Scaling vectors 尺度向量 504
Scanning electron microscope (SEM) 扫描电子显微镜 45, 137, 164, 278
Scotopic vision 暗视觉 59
Segmentation 分割 711–816
color 彩色~ 465–472

definition ～的定义 712
edge-based 基于边缘的～ 见 *Edge detection*
foundation 基础 712–717
frequency-based 基于频率的～ 804–807
line detection 线检测 719
motion and 运动和～ 800–807
point detection 点检测 718
region growing 区域生长 见 *Region-based segmentation*
texture based 基于～的纹理 791
thresholding ～阈值处理 见 *Thresholding*
watersheds 分水岭 见 *Watersheds*
Sensors 传感器 50, 68–74, 76
acquisition and 获取与～ 68–74
arrays ～阵列 72, 76
cooling 冷却～ 98
image formation model for ～的图像形成模型 73–74
imaging component for ～成像成分 50
sampling and quantization using 使用～取样和量化 76
single 单～ 70
strips ～条带 70–72, 76
Sequential baseline system 顺序基线系统 602
Series expansions 级数展开 499–501, 508–510
Set operations 集合操作(运算) 102–105, 106–107, 650–652, 652–657, 657–661 见 *Fuzzy sets*
basics of ～的基础 102–105
closing ～的开操作 657–391
crisp 干脆～ 106
dilation ～膨胀 655–657
erosion ～腐蚀 652–655, 657
fuzzy concept of ～的模糊概念 106–107, 195–213
morphological image processing and 形态学图像处理和～ 650–652, 652–657, 657–661
opening 开操作 657–661, 690–692
Shading correction 阴影校正 100–101, 695, 763, 783
Shannon's first theorem 山农第一定理 555–556
Shape numbers 形状数 838–839, 925–926
Sharpening (尖)锐化 见 *Filtering*
Shepp-Logan phantom Shepp-Logan 幻影 394
Shrinking 缩小 见 *Image resampling*
Sifting property 取样特性 见 *Impulse*
Signal-to-noise (SNR) ratios 信噪比 376, 557
Signatures 标记图 830–832
Simultaneous contrast 同时对比 63–64
Single-pixel operations 单像素操作 107
Skeletons 骨架 673–676, 834–837
Slope overload 斜率过载 620
Smoothing 平滑 见 *Filtering*
SMPTE 活动图像和电视工程师协会 560
Sobel gradient operators Sobel 梯度算子 见 *Spatial filters*
Software for imaging 图像处理软件 51–52
Spatial coordinates 空间坐标 23, 77
Spatial domain 空间域
convolution ～卷积 见 *Convolution*
correlation ～相关 见 *Correlation*
definition ～的定义 77
image transform difference ～图像变换差 115–116
filtering ～滤波 见 *Spatial filtering*
frequency domain 频率域与～
correspondence ～对应 285
operations ～操作 107–114
Spatial filters 空间滤波器 见 *Spatial filtering*
adaptive local 自适应局部～ 352–354
adaptive median 自适应中值～ 354–357
alpha-trimmed 修正的阿尔法～ 349
arithmetic mean 算术平均～ 344
averaging 平均～ 174
contraharmonic mean 反调和平均～ 345
defined ～的定义 128
generating ～生成 173
geometric mean 几何平均～ 345
gradient ～梯度 187
harmonic mean 调和平均～ 345
highboost 高提升～ 184
isotropic 各向同性～ 182
Laplacian 拉普拉斯～ 182–185
lowpass 低通～ 174
max 最大～ 179, 348
median 中值～ 178, 348
midpoint 中点～ 349
min 最小～ 179, 348
order statistic 统计排序～ 178–179, 347
Roberts 罗伯特～ 189
sharpening (尖)锐化～ 179–190
smoothing 平滑～ 174–179, 344
Sobel Sobel～ 189
unsharp mask 非锐化模板～ 184
vector representation ～向量表示 172
weighted average 加权平均～ 175
Spatial filtering 空间滤波 126–220, 344–357
adaptive local 自适应局部～ 352–354
adaptive median 自适应中值～ 354–357
convolution and 卷积和～ 168–172
correlation and 相关和～ 168–172
defined ～的定义 128
enhancement methods combined 组合～的增强方法 191–195
fundamentals of ～的基础 166–174
fuzzy techniques for ～模糊技术 195–213
linear 线性～ 167–177
masks ～模板 见 *Spatial filters*
mechanics of ～的机理 167
noise reduction by ～降噪 344–357
nonlinear 非线性～ 167, 177–179, 344–357
order-statistic 统计排序～ 177–179, 347
sharpening (尖)锐化～ 179–190
smoothing 平滑～ 174–179
vector representation of ～的向量表示 172–173
Spatial operations 空间操作 107–114
Spatial redundancy 空间冗余 549, 551–552
Spatial resolution 空间分辨率 81–87
Spatial techniques for motion in segmentation 分割中运动的空间技术 800–804
Spatial variables 空间变量 77
Spectrum 谱 见 *Fourier transform, Discrete Fourier transform*
Standard definition (SD) television 标准清晰度电视 547–548
Standards for image compression 图像压缩标准 560–562, 563
Statistical moments 统计矩 见 *Moments*
Stochastic image processing 随机图像处理 120
Storage capacity for imaging 成像的存储容量 52
String descriptions 串描述 886–887, 926–928
Subband coding 子带编码 488–495
Subjective brightness 主观亮度 61
Subsampling pyramids 子取样金字塔 486
Subspace analysis trees 子空间分析树 533
Successive doubling 逐次加倍 322
Sum of absolute distortions (SAD) 绝对失真之和 612
Superposition integral 超位置积分 367
Super-sampling 超取样 253
Symbol coders 符号编码器 559
Symbol-based coding 基于符号的编码 581–584
Symlets 对称小波 527–529
Synthesis filter banks 综合滤波器组 492, 521–522, 525–526
Synthetic imaging 合成成像 46–47

T

Temporal redundancy 空间冗余 549, 551–552

Texture 纹理 697–698, 791, 849–861
co-occurrence matrix for ~的共生矩阵 852–858
description by ~描述 849–861
gray-scale morphology and 灰度形态学和~ 697–698
intensity histogram for ~的灰度直方图 850–852
segmentation ~分割 697–698, 791
spectral approaches to ~的谱方法 859–861
statistical approaches for ~的统计方法 850–858
structural approaches for ~的结构方法 858–859
Thematic bands 热波段 36
Thickening 粗化 见 *Morphological image processing*
Thinning 细化 见 *Morphological image processing*
Threshold 阈值(处理) 见 *Thresholding*
basic ~基础 129, 760, 763
Bayes 贝叶斯~ 764, 897, 903–904
coding ~编码 597–601
color ~彩色 467
combined with blurring ~与模糊相组合 191
combined with gradient ~与梯度相组合 735, 771
combined with Laplacian ~与拉普拉斯相组合 772
global 全局~ 763
hysteresis 滞后~ 744, 776
local 局部~ 780–783
optimum 最佳~ 764
Otsu Otsu~ 695, 764, 774
multiple 多~ 744, 761, 774–778
multivariable 多变量~ 467, 783–785
variable 可变~ 778
Thresholding 阈值处理 129, 137, 530, 599–601, 735–736, 760–785
basics ~基础 760
Bayes 贝叶斯~ 764, 897, 903–904
coding implementation ~编码实现 599–601
edges using in ~中使用的边缘 771
function ~函数 129, 137
global 全局~ 760, 763–778
gradients, combined with ~梯度，与~的组合 735–736
hard 硬~ 530
illumination and 照度和~ 762–763
intensity ~灰度 760–761
Laplacian, combined with 拉普拉斯，与~相组合 772
local 局部~ 780–783
measure of separability 可分离性度量~ 767
moving averages 移动平均~ 781
multiple thresholds 多阈值~ 774
multivariable 多变量~ 467, 783–785
noise in ~中的噪声 761–762
object point for ~的目标点 760
optimum 最佳~ 764
Otsu Otsu~ 695, 764, 774
reflectance and 反射和~ 762–763
segmentation and 分割和~ 760–785
smoothing in ~中的平滑 769
soft 软~ 530
variable ~变量 760, 778–785
Tie (control) points 约束(控制)点 112
TIFF 一种图像压缩格式 560, 563, 573
Tight frame 紧框架 501
Tiling images 贴瓷砖图像 46, 523
Time-frequency tiles (or plane) 时间-频率片(或平面) 522–523
Tokens 标记 582
Top-hat transformation 顶帽变换 694–696
Top-hat by reconstruction 顶帽重建 699
Topological descriptors 拓扑描绘子 845–849
Transformation 变换 109–114, 126–220, 662–663, 694–696
Affine 仿射~ 109–111
bottom-hat 底帽~ 694–696
domains in ~中的域 126
geometric (rubber-sheet) 几何(橡皮膜)~ 见 *Geometric transformations*
gray-scale morphology and 灰度级形态学和~ 694–696
hit-or-miss 击中或击不中~ 662–663
intensity 灰度~ 126–220
kernels ~核 117
morphological image processing and 形态学图像处理和~ 662–663
rubber sheet 橡皮膜~ 109, 845
spatial 空间~ 107, 127–193
top-hat 顶帽~ 694–696
top-hat by reconstruction 顶帽重建 699
Transforms 变换 115–118, 126, 388, 390–395, 496–499, 508–515, 523–532, 588–606
block transform coding 块变换编码~ 588–606
discrete cosine 离散余弦~ 591
domains in ~中的域 115–116, 126
discrete cosine 离散余弦~ 118, 561, 591
discrete Karhunen-Loeve 离散 Karhunen-Loeve~ 867
Fourier 傅里叶~ 见 *Fourier transform*
Haar 哈尔~ 118, 496–499
Hotelling 霍特林~ 867–874
Hough 霍夫~ 见 *Hough transform*
image (2-D linear) ~图像(二维线性) 115–118
morphological 形态学~ 见 *Morphological image processing*
pair ~对 116
principal components 主成分~ 864–874
Radon 雷登~ 388, 390–395
selection of for block transform 为块变换选取的~
coding ~编码 589–595
slant 斜~ 118
Walsh-Hadamard 沃尔什-哈达玛~ 118, 590
wavelet 小波~ 508–515, 523–532 见 *Wavelets*
Transmission electron microscope (TEM) 发射电子显微镜 45
Trichromatic coefficients 三色系数 421

U
Ultra large scale integration (ULSI) 超大规模集成 27
Ultrasound imaging 超声波成像 42, 68, 390, 410
Ultraviolet 紫外线 29, 33, 59, 66, 67
Unary codes 一元编码 566
Unbiased estimate 无偏估计 163
Uniform 均匀 见 Noise
Unit delays 单位延迟 488–489
Unit discrete impulse 单位离散冲激 见 *Impulse*
Unit impulse 单位冲激 见 *Impulse*
Units of measurement 单位度量 66, 67, 80–82
bits for image storage 图像存储的比特 80–82
electromagnetic (EM) spectrum 电磁波谱 66, 67
intensity resolution 灰度分辨率 81–82
spatial resolution 空间分辨率 81
Unsharp masking 非锐化模板 184–187, 310–311
Upsampling 上采(取)样 486–487

V
Variable thresholding 可变阈值处理 见 *Thresholding*
Variable-length code 变长编码 551, 564–566
Variance of intensity 灰度的方差 见 *Moments*
VC-1 compression VC-1 压缩 560, 563, 616
Vector operations 向量操作 114–115, 172–173, 446–448
full-color image processing 全彩色图像处理~ 446–448
matrix operations and 矩阵操作和~ 114–115
spatial filtering 空间滤波~ 172–173
Very large scale integration (VLSI) 甚大规模集成电路 27
Visible band of the EM spectrum 电磁波谱的可见波段 34–40, 66–67
Visible watermarks 可见水印 637
Vision 视觉 见 *Visual perception*
human 人类~ 58–65, 418, 740, 800
machine 机器~ 24–25, 28, 928

Visual perception　视觉感知　58–65, 417–423
absorption of light　光的吸收　418–419
brightness adaptation　亮度适应　61–65
color image processing and　彩色图像处理和～　417–423
discrimination between changes　变化间的分辨～　58–65r
human eye physical structure　人眼的自然结构　58–60
image formation in eye　眼中的图像形成　60–61
Mach bands　马赫带　63, 64
optical illusions　错觉　64–65
simultaneous contrast　同时对比　63–64
subjective brightness　主观亮度　61–62
Weber ratio　韦伯比　62–63

W

Walsh-Hadamard transform (WHT)　活尔什-哈达玛变换　590–591
Watermarking digital images　水印数字图像　636–643
block diagram for　～的方框图　639
reasons for　～的原因　636
Watermarks　水印　636–643
attacks on　对～的攻击　642–643
fragile invisible　脆弱不可见～　639
insertion and extraction　～插入和提取　637–638, 640–642
invisible watermark　不可见～　638
private (or restricted key)　私钥（或受限密钥）～　639
public (or unrestricted key)　公钥（或不受限密钥）～　639
robust invisible　鲁棒不可见～　639
visible watermark　可见～　637
Watersheds (morphological)　分水岭（形态学）～　791–800
algorithm for　～算法　796–798
dam construction for　～水坝构建　794–796
knowledge incorporation in　～中的知识合并　791
markers used for　～使用的标记　798–800
segmentation using　使用～的分割　791–800
Wavelet coding　小波编码　626–636
decomposition level selection　～的分解级数选取　628–629
JPEG-2000 compression　JPEG-2000 压缩　629–636
quantizer design for　～的量化器设计　629
selection of wavelets for　小波的选取　626–628
Wavelet functions　小波函数　505
coefficients of　～的系数　506
Haar　哈尔～　506–507
separable 2D　可分二维～　524
time-frequency characteristics　～的时间-频率特性　522–523
Wavelet vectors　小波向量　506
Wavelet packets　小波包　532–541
binary tree representation　～的二叉树表示　532–540
cost functions for choosing　选择～的代价函数　537–540
subspace analysis tree　～的子空间分析树　533
Wavelets　小波　49, 483–546
compression　～压缩　626–629
continuous wavelet transform (CWT)　连续小波变换　513–515
discrete wavelet transform (DWT)　离散小波变换　510–512, 524
edge detection　边缘检测　529–530
fast wavelet transform (FWT)　快速小波变换　515–523, 524–527, 532–541
functions　～函数　505–508
JPEG-2000　一种压缩格式　629–635
Mexican hat　墨西哥草帽～　514–515
multiresolution processing and　多分辨率处理　483–546
noise removal　噪声消除　530–532
one-dimensional transforms　一维～变换　508–515
packets　～包　532–541
series expansions　～级数展开　508–510
transforms　～变换　508–515, 523–532
two-dimensional transforms　～二维变换　523–532
Weber ratio　韦伯比　62–63
Weighting function　加权函数　363
White noise　白噪声　见 *Noise*
Wiener filtering　维纳滤波　374–379
WMV9 compression　WMV9 压缩　560, 563, 616

X

X-rays　X 射线　31, 137, 179, 311, 346, 384, 385, 387, 439, 442, 668, 689, 693, 719, 753, 786, 790

Z

Zero crossing property　零交叉特性　182, 725, 736–739
Zero-memory source　零记忆信源　554
Zero-phase-shift filters　零相移滤波器　284, 316
Zonal coding implementation　区域编码实现　598–599
Zooming　缩放　见 *Image zooming*